

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/45748 A1

(51) International Patent Classification⁷: **A61K 48/00, C12Q 1/70**

NJ 07065-0907 (US). FU, Tong-Ming [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(21) International Application Number: **PCT/US00/34724**

(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(22) International Filing Date:
21 December 2000 (21.12.2000)

(81) Designated States (*national*): AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CI, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, IU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PI, PT, RO, RU, SD, SE, SG, SI, SK, SI, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

Published:

— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). PERRY, Helen, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). CASIMIRO, Danilo, R. [PH/US]; 126 East Lincoln Avenue, Rahway,**



WO 01/45748 A1

(54) Title: POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1 POL AND MODIFIED HIV-1 POL

(57) Abstract: Pharmaceutical compositions which comprise HIV Pol DNA vaccines are disclosed, along with the production and use of these DNA vaccines. The pol-based DNA vaccines of the invention are administered directly introduced into living vertebrate tissue, preferably humans, and preferably express inactivated versions of the HIV Pol protein devoid of protease, reverse transcriptase activity, RNase H activity and integrase activity, inducing a cellular immune response which specifically recognizes human immunodeficiency virus-1 (HIV-1). The DNA molecules which comprise the open reading frame of these DNA vaccines are synthetic DNA molecules encoding codon optimized HIV-1 Pol and codon optimized inactive derivatives of optimized HIV-1 Pol, including DNA molecules which encode inactive Pol proteins which comprise an amino terminal leader peptide.

TITLE OF THE INVENTION**POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1****5 POL AND MODIFIED HIV-1 POL****CROSS-REFERENCE TO RELATED APPLICATIONS**

This application claims the benefit, under 35 U.S.C. §119(e), of U.S. provisional application 60/171,542, filed December 22, 1999.

10

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

Not Applicable

15 REFERENCE TO MICROFICHE APPENDIX

Not Applicable

FIELD OF THE INVENTION

The present invention relates to HIV Pol polynucleotide pharmaceutical products, as well as the production and use thereof which, when directly introduced into living vertebrate tissue, preferably a mammalian host such as a human or a non-human mammal of commercial or domestic veterinary importance, express the HIV Pol protein or biologically relevant portions thereof within the animal, inducing a cellular immune response which specifically recognizes human immunodeficiency virus-1 (HIV-1). The polynucleotides of the present invention are synthetic DNA molecules encoding codon optimized HIV-1 Pol and derivatives of optimized HIV-1 Pol, including constructs wherein protease, reverse transcriptase, RNase H and integrase activity of HIV-1 Pol is inactivated. The polynucleotide vaccines of the present invention should offer a prophylactic advantage to previously uninfected individuals and/or provide a therapeutic effect by reducing viral load levels within an infected individual, thus prolonging the asymptomatic phase of HIV-1 infection.

BACKGROUND OF THE INVENTION

- Human Immunodeficiency Virus-1 (HIV-1) is the etiological agent of acquired human immune deficiency syndrome (AIDS) and related disorders. HIV-1 is an RNA virus of the Retroviridae family and exhibits the 5'LTR-gag-pol-env-LTR 3' organization of all retroviruses. The integrated form of HIV-1, known as the provirus, is approximately 9.8 Kb in length. Each end of the viral genome contains flanking sequences known as long terminal repeats (LTRs). The HIV genes encode at least nine proteins and are divided into three classes; the major structural proteins (Gag, Pol, and Env), the regulatory proteins (Tat and Rev); and the accessory proteins (Vpu, Vpr, Vif and Nef).
- The *gag* gene encodes a 55-kilodalton (kDa) precursor protein (p55) which is expressed from the unspliced viral mRNA and is proteolytically processed by the HIV protease, a product of the *pol* gene. The mature p55 protein products are p17 (matrix), p24 (capsid), p9 (nucleocapsid) and p6.
- The *pol* gene encodes proteins necessary for virus replication; a reverse transcriptase, a protease, integrase and RNase H. These viral proteins are expressed as a Gag-Pol fusion protein, a 160 kDa precursor protein which is generated via a ribosomal frame shifting. The viral encoded protease proteolytically cleaves the Pol polypeptide away from the Gag-Pol fusion and further cleaves the Pol polypeptide to the mature proteins which provide protease (Pro, P10), reverse transcriptase (RT, P50), integrase (IN, p31) and RNase H (RNase, p15) activities.
- The *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating CD4 expression, disturbing T-cell activation and stimulating HIV infectivity.
- The *env* gene encodes the viral envelope glycoprotein that is translated as a 160-kilodalton (kDa) precursor (gp160) and then cleaved by a cellular protease to yield the external 120-kDa envelope glycoprotein (gp120) and the transmembrane 41-kDa envelope glycoprotein (gp41). Gp120 and gp41 remain associated and are displayed on the viral particles and the surface of HIV-infected cells.
- The *tat* gene encodes a long form and a short form of the Tat protein, a RNA binding protein which is a transcriptional transactivator essential for HIV-1 replication.
- The *rev* gene encodes the 13 kDa Rev protein, a RNA binding protein. The Rev protein binds to a region of the viral RNA termed the Rev response element

(RRE). The Rev protein is promotes transfer of unspliced viral RNA from the nucleus to the cytoplasm. The Rev protein is required for HIV late gene expression and in turn, HIV replication.

Gp120 binds to the CD4/chemokine receptor present on the surface of helper T-lymphocytes, macrophages and other target cells in addition to other co-receptor molecules. X4 (macrophage tropic) virus show tropism for CD4/CXCR4 complexes while a R5 (T-cell line tropic) virus interacts with a CD4/CCR5 receptor complex. After gp120 binds to CD4, gp41 mediates the fusion event responsible for virus entry. The virus fuses with and enters the target cell, followed by reverse transcription of its single stranded RNA genome into the double-stranded DNA via a RNA dependent DNA polymerase. The viral DNA, known as provirus, enters the cell nucleus, where the viral DNA directs the production of new viral RNA within the nucleus, expression of early and late HIV viral proteins, and subsequently the production and cellular release of new virus particles. Recent advances in the ability to detect viral load within the host shows that the primary infection results in an extremely high generation and tissue distribution of the virus, followed by a steady state level of virus (albeit through a continual viral production and turnover during this phase), leading ultimately to another burst of virus load which leads to the onset of clinical AIDS. Productively infected cells have a half life of several days, whereas chronically or latently infected cells have a 3-week half life, followed by non-productively infected cells which have a long half life (over 100 days) but do not significantly contribute to day to day viral loads seen throughout the course of disease.

Destruction of CD4 helper T lymphocytes, which are critical to immune defense, is a major cause of the progressive immune dysfunction that is the hallmark of HIV infection. The loss of CD4 T-cells seriously impairs the body's ability to fight most invaders, but it has a particularly severe impact on the defenses against viruses, fungi, parasites and certain bacteria, including mycobacteria.

Effective treatment regimens for HIV-1 infected individuals have become available recently. However, these drugs will not have a significant impact on the disease in many parts of the world and they will have a minimal impact in halting the spread of infection within the human population. As is true of many other infectious diseases, a significant epidemiologic impact on the spread of HIV-1 infection will only occur subsequent to the development and introduction of an effective vaccine. There are a number of factors that have contributed to the lack of successful vaccine

development to date. As noted above, it is now apparent that in a chronically infected person there exists constant virus production in spite of the presence of anti-HIV-1 humoral and cellular immune responses and destruction of virally infected cells. As in the case of other infectious diseases, the outcome of disease is the result of a
5 balance between the kinetics and the magnitude of the immune response and the pathogen replicative rate and accessibility to the immune response. Pre-existing immunity may be more successful with an acute infection than an evolving immune response can be with an established infection. A second factor is the considerable genetic variability of the virus. Although anti-HIV-1 antibodies exist that can
10 neutralize HIV-1 infectivity in cell culture, these antibodies are generally virus isolate-specific in their activity. It has proven impossible to define serological groupings of HIV-1 using traditional methods. Rather, the virus seems to define a serological "continuum" so that individual neutralizing antibody responses, at best, are effective against only a handful of viral variants. Given this latter observation, it
15 would be useful to identify immunogens and related delivery technologies that are likely to elicit anti-HIV-1 cellular immune responses. It is known that in order to generate CTL responses antigen must be synthesized within or introduced into cells, subsequently processed into small peptides by the proteasome complex, and translocated into the endoplasmic reticulum/Golgi complex secretory pathway for
20 eventual association with major histocompatibility complex (MHC) class I proteins. CD8⁺ T lymphocytes recognize antigen in association with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein. Activation of naive CD8⁺ T cells into activated effector or memory cells generally requires both TCR engagement of antigen as described above as well as engagement of costimulatory proteins. Optimal
25 induction of CTL responses usually requires "help" in the form of cytokines from CD4⁺ T lymphocytes which recognize antigen associated with MHC class II molecules via TCR and CD4 engagement.

Larder, et al., (1987, *Nature* 327: 716-717) and Larder, et al., (1989, *Proc. Natl. Acad. Sci.* 86: 4803-4807) disclose site specific mutagenesis of HIV-1 RT and
30 the effect such changes have on *in vitro* activity and infectivity related to interaction with known inhibitors of RT.

Davies, et al. (1991, *Science* 252: 88-95) disclose the crystal structure of the RNase H domain of HIV-1 Pol.

Schatz, et al. (1989, *FEBS Lett.* 257: 311-314) disclose that mutations Glu478Gln and His539Phe in a complete HIV-1 RT/RNase H DNA fragment results in defective RNase activity without effecting RT activity.

5 Mizrahi, et al. (1990, *Nucl. Acids. Res.* 18: pp. 5359-5353) disclose additional mutations Asp443Asn and Asp498Asn in the RNase region of the *pol* gene which also results in defective RNase activity. The authors note that the Asp498Asn mutant was difficult to characterize due to instability of this mutant protein.

Leavitt, et al. (1993, *J. Biol. Chem.* 268: 2113-2119) disclose several 10 mutations, including a Asp64Val mutation, which show differing effect on HIV-1 integrase (IN) activity.

Wiskerchen, et al. (1995, *J. Virol.* 69: 376-386) disclose singe and double mutants, including mutation of aspartic acid residues which effect HIV-1 IN and viral replication functions.

It would be of great import in the battle against AIDS to produce a 15 prophylactic- and/or therapeutic-based HIV vaccine which generates a strong cellular immune response against an HIV infection. The present invention addresses and meets this needs by disclosing a class of DNA vaccines based on host delivery and expression of modified versions of the HIV-1 gene, *pol*.

20 SUMMARY OF THE INVENTION

The present invention relates to synthetic DNA molecules (also referred to herein as "polynucleotides") and associated DNA vaccines (also referred to herein as "polynucleotide vaccines") which elicit cellular immune and humoral responses upon administration to the host, including primates and especially humans, and also 25 including a non-human mammal of commercial or domestic veterinary importance. An effect of the cellular immune-directed vaccines of the present invention should be the lower transmission rate to previously uninfected individuals and/or reduction in the levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to 30 DNA vaccines which encode various forms of HIV-1 Pol, wherein administration, intracellular delivery and expression of the HIV-1 Pol gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized versions of wild type HIV-1 Pol, codon optimized versions of HIV-1 Pol fusion proteins, and codon optimized versions of HIV-1 Pol

proteins and fusion protein, including but not limited to *pol* modifications involving residues within the catalytic regions responsible for RT, RNase and IN activity within the host cell.

- A particular embodiment of the present invention relates to codon optimized wt-pol DNA constructs wherein DNA sequences encoding the protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase H activity) and IN integrase activity. The nucleotide sequence of a DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1 and the corresponding amino acid sequence of the expressed protein is disclosed herein as SEQ ID NO:2.

The present invention preferably relates to a HIV-1 DNA pol construct which is devoid of DNA sequences encoding any PR activity, as well as containing a mutation(s) which at least partially, and preferably substantially, abolishes RT, RNase and/or IN activity. One type of HIV-1 pol mutant may include but is not limited to a mutated DNA molecule comprising at least one nucleotide substitution which results in a point mutation which effectively alters an active site within the RT, RNase and/or IN regions of the expressed protein, resulting in at least substantially decreased enzymatic activity for the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct in a DNA molecule which contains at least one point mutation which alters the active site of the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. To this end, an especially preferred HIV-1 DNA pol construct is exemplified herein and contains nine codon substitution mutations which results in an inactivated Pol protein (IA Pol: SEQ ID NO:4, Figure 2A-C) which has no PR, RT, RNase or IN activity, wherein three such point mutations reside within each of the RT, RNase and IN catalytic domains. Any combination of the mutations disclosed herein may suitable and therefore may be utilized as an IA-Pol-based vaccine of the present invention. While addition and deletion mutations are contemplated and within the scope of the invention, the

preferred mutation is a point mutation resulting in a substitution of the wild type amino acid with an alternative amino acid residue.

Another aspect of the present invention is to generate HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as 5 the leader peptide from human tPA. To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame disclosed herein as SEQ ID NO:6.

10 The present invention especially relates to a HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) which comprises a leader peptide, such as the human tPA leader, at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. Any such HIV-1 DNA pol mutant disclosed in the above paragraphs is suitable for fusion downstream of a leader 15 peptide, including but by no means limited to the human tPA leader sequence. Therefore, any such leader peptide-based HIV-1 pol mutant construct may include but is not limited to a mutated DNA molecule which effectively alters the catalytic activity of the RT, RNase and/or IN region of the expressed protein, resulting in at least substantially decreased enzymatic activity one or more of the RT, RNase H and/or IN functions of 20 HIV-1 Pol. In a preferred embodiment of this portion of the invention, a leader peptide/HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct is a DNA molecule which contains at least one point mutation which alters the active site and catalytic activity within the RT, RNase H and IN 25 domains of Pol, such that each activity is at least substantially abolished, and preferably totally abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. An especially preferred embodiment of this portion of the invention relates to a human tPA leader fused to the IA-Pol protein comprising the nine mutations shown 30 in Table 1. The DNA molecule is disclosed herein as SEQ ID NO:7 and the expressed tPA-IA Pol protein comprises a fusion junction as shown in Figure 3. The complete amino acid sequence of the expressed protein is set forth in SEQ ID NO:8.

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified

proteins set forth below as SEQ ID NOs: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

The present invention also relates to non-codon optimized versions of DNA molecules and associated polynucleotides and associated DNA vaccines which encode the various wild type and modified forms of the HIV Pol protein disclosed herein. Partial or fully codon optimized DNA vaccine expression vector constructs are preferred, but it is within the scope of the present invention to utilize "non-codon optimized" versions of the constructs disclosed herein, especially modified versions of HIV Pol which are shown to promote a substantial cellular immune and humoral immune responses subsequent to host administration.

The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors utilized in the present invention include but are not limited to constructs which comprise the cytomegalovirus promoter with the intron A sequence (CMV-intA) and a bovine growth hormone transcription termination sequence. In addition, DNA plasmid vectors of the present invention preferably comprise an antibiotic resistance marker, including but not limited to an ampicillin resistance gene, a neomycin resistance gene or any other pharmaceutically acceptable antibiotic resistance marker. In addition, an appropriate polylinker cloning site and a prokaryotic origin of replication sequence are also preferred. Specific DNA vectors exemplified herein include V1, V1J (SEQ ID NO:13), V1Jneo (SEQ ID NO:14), V1Jns (Figure 1A, SEQ ID NO:15), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding codon optimized HIV-1 Pol (e.g. SEQ ID NO:2), codon optimized HIV-1 Pol fused to an amino terminal localized leader sequence (e.g. SEQ ID NO:6), and especially preferable, and the essence of the present invention, biologically inactive Pol proteins (IA Pol; e.g., SEQ ID NO:4) devoid of significant PR, RT, RNase or IN

activity associated with wild type Pol and a concomitant construct which contains a leader peptide at the amino terminal region of the IA Pol protein. These constructs are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. Preferred DNA vaccines of the
5 present invention comprise codon optimized DNA molecules encoding codon optimized HIV-1 Pol and inactivated version of Pol, ligated in DNA vectors disclosed herein, or any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B
10 and SEQ ID NO:28.

Therefore, the present invention relates to DNA vaccines which include, but are in no way limited to V1Jns-WTPol (comprising the DNA molecule encoding WT Pol, as set forth in SEQ ID NO:2), V1Jns-tPA-WTPol, (comprising the DNA molecule encoding tPA Pol, as set forth in SEQ ID NO:6), V1Jns-IAPol (comprising 15 the DNA molecule encoding IA Pol, as set forth in SEQ ID NO:4), and V1Jns-tPA-IAPol, (comprising the DNA molecule encoding tPA-IA Pol, as set forth in SEQ ID NO:8). Especially preferred are V1Jns-IAPol and V1Jns-tPA-IAPol, as exemplified in Example Section 2.

The present invention also relates to HIV Pol polynucleotide
20 pharmaceutical products, as well as the production and use thereof, wherein the DNA vaccines are formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention, namely by promoting an enhanced cellular and/or humoral response subsequent to inoculation. A preferred adjuvant is an aluminum phosphate-based adjuvant or a
25 calcium phosphate based adjuvant, with an aluminum phosphate adjuvant being especially preferred. Another preferred adjuvant is a non-ionic block copolymer, preferably comprising the blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. These adjuvanted forms comprising the DNA vaccines disclosed herein are useful in
30 increasing cellular responses to DNA vaccination.

As used herein, a DNA vaccine or DNA polynucleotide vaccine is a DNA molecule (i.e., "nucleic acid", "polynucleotide") which contains essential regulatory elements such that upon introduction into a living, vertebrate cell, it is able to direct the cellular machinery to produce translation products encoded by the respective pol

genes of the present invention.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A-B shows schematic representation of DNA vaccine expression vectors V1Jns (A) and V1Jns-tPA (B) utilized for HIV-1 pol and HIV-1 modified pol constructs.

Figure 2A-C shows the nucleotide (SEQ ID NO:3) and amino acid sequence (SEQ ID NO:4) of IA-Pol. Underlined codons and amino acids denote mutations, as listed in Table 1.

Figure 3 shows the codon optimized nucleotide and amino acid sequences through the fusion junction of tPA-IA-Pol (contained within SEQ ID NOs: 7 and 8, respectively). The underlined portion represents the NH₂-terminal region of IA-Pol.

Figure 4 shows generation of a humoral response (measured as the geometric means of anti-RT endpoint titers) from mice immunized with one or two doses of codon optimized V1Jns-IApol and V1Jns-tpa-IApol. A portion of mice that received 30 ug of each plasmid was boosted at T=8 wks; sera from all mice were collected at 4 wk post dose 2.

Figure 5 shows the number of IFN-gamma secreting cells per 10e6 cells following stimulation with pools of either CD4⁺ (aa641-660, aa731-750) or CD8⁺ (aa201-220, aa311-330, aa571-590, aa781-800) specific peptides of splenocytes (pool of 5 spleens/cohort) from control mice and those vaccinated with increasing single dose of codon optimized V1Jns-IApol or 30 ug of codon optimized V1Jns-tpa-IApol (13 wks post dose 1). Mice (n=5) vaccinated with a second dose of 30 ug of either plasmid were analyzed in an Elispot assay at 6 wks post dose 2. Reported are the sums of the number of spots stimulated by each individual CD8⁺ peptides because the spots in the wells to which the pool was added are too dense to acquire accurate counts. The CD4⁺ cell counts are taken from the responses to the peptide pool. Error bars represent standard deviations for counts from triplicate wells per sample per antigen.

Figure 6A-C shows ELIspot analysis of peripheral blood cells collected from rhesus macaques immunized three times (T=0, 4, 8 wks) with 5 mgs of codon optimized HIV-1 Pol expressing plasmids. Antigen-specific IFN-gamma secretion was stimulated by adding one of two pools consisting of 20-mer peptides derived from vaccine sequence (mpol-1, aa1-420; mpol-2, aa411-850). (A) Frequencies of

spot-forming cells (SFC) as a function of time for 3 monkeys (Tag No. 94R008, 94R013, 94R033) vaccinated with V1Jns-IApol. The reported values are corrected for background responses without peptide restimulation. (B) Frequencies of spot-forming cells (SFC) as a function of time for 3 monkeys (Tag No. 920078, 920073, 5 94R028) vaccinated with 5mgs of V1Jns-tpa-IApol. (C) ELIspot responses were also measured from a monkey (920072) that did not receive any immunization.

Figure 7A-B show bulk CTL killing from rhesus macaques immunized with codon optimized V1Jns-IApol (A) or codon optimized V1Jns-tpa-IApol (B) at 8 weeks following the third vaccination. Restimulation was performed using recombinant 10 vaccinia virus expressing pol and target cells were prepared by pulsing with the peptide pools, mpol-1 and mpol-2.

Figure 8 shows detection of *in vitro* pol expression from cell lysates of 293 cells transfected with 10 ug of various pol constructs. Bands were detected using anti-serum from an HIV-1 seropositive human subject. Equal amounts of total protein 15 were loaded for each lane. The lanes contain the lysates from cells transfected with the following: 1: mock; 2: V1Jns-wt-pol; 3: V1Jns-IApol (codon optimized); 4: V1Jns-tpa-IApol (codon optimized); 5: V1Jns-tpa-pol (codon optimized); 6: V1R-wt-pol (codon optimized); 7: blank; and 8: 80 ng RT.

Figure 9 shows the geometric mean anti-RT titers (GMT) plus the standard 20 errors of the geometric means for cohorts of 5 mice that received one (open circles) or two doses (solid circles) of 1, 10, 100 μ g of V1R-wt-pol (codon optimized) or V1Jns-wt-pol. Sera from all animals were collected at 2 weeks post dose 2 (or 7 wks post dose 1) and assayed simultaneously. Statistical analyses were performed to compare cohorts that received the same amount and number of immunization of either 25 plasmids; p values (two-tail) less than 5% are above the bars that connect the correlated cohorts to reflect statistically significant differences.

Figure 10 shows cellular immune responses in BALB/c mice vaccinated i.m. with 1 (pd1) or 2 (pd2) doses of varying amounts of either wt-pol (virus derived) or 30 wt-pol (codon optimized) plasmids. At 3 wks post dose 2, frequencies of IFN- γ - secreting splenocytes are determined from pools of 5 spleens per cohort against mixtures of either CD4 $^{+}$ peptides (aa21-40, aa411-430, aa531-550, aa641-660, aa731-750, aa771-790) or CD8 $^{+}$ peptides (aa201-220, aa311-330) at 4 μ g/mL final concentration per peptide.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to synthetic DNA molecules and associated DNA vaccines which elicit CTL and Th cellular immune responses upon administration to the host, including primates and especially humans. An effect of the 5 cellular immune-directed vaccines of the present invention should be a lower transmission rate to previously uninfected individuals and/or reduction in the levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to DNA vaccines which encode various forms of HIV-1 Pol, wherein administration, intracellular 10 delivery and expression of the HIV-1 Pol gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized wild type Pol (without Pro activity) and various codon optimized inactivated HIV-1 Pol proteins. The HIV-1 *pol* constructs disclosed herein are especially preferred for pharmaceutical uses, especially for human administration as a 15 DNA vaccine. The HIV-1 genome employs predominantly uncommon codons compared to highly expressed human genes. Therefore, the pol open reading frame has been synthetically manipulated using optimal codons for human expression. As noted above, a preferred embodiment of the present invention relates to DNA molecules which comprise a HIV-1 pol open reading frame, whether encoding full 20 length pol or a modification or fusion as described herein, wherein the codon usage has been optimized for expression in a mammal, especially a human.

The synthetic *pol* gene disclosed herein comprises the coding sequences for the reverse transcriptase (or RT which consists of a polymerase and RNase H activity) and integrase (IN). The protein sequence is based on that of Hxb2r, a clonal isolate of 25 IIIB; this sequence has been shown to be closest to the consensus clade B sequence with only 16 nonidentical residues out of 848 (Korber, et al., 1998, Human retroviruses and AIDS, Los Alamos National Laboratory, Los Alamos, New Mexico). The skilled artisan will understand after review of this specification that any available HIV-1 or HIV-2 strain provides a potential template for the generation of HIV pol 30 DNA vaccine constructs disclosed herein. It is further noted that the protease gene is excluded from the DNA vaccine constructs of the present invention to insure safety from any residual protease activity in spite of mutational inactivation. The design of the gene sequences for both wild-type (wt-pol) and inactivated pol (IA-pol) incorporates the use of human preferred ("humanized") codons for each amino acid

residue in the sequence in order to maximize *in vivo* mammalian expression (Lathe, 1985, J. Mol. Biol. 183:1-12). As can be discerned by inspecting the codon usage in SEQ ID NOs: 1, 3, 5 and 7, the following codon usage for mammalian optimization is preferred: Met (ATG), Gly (GGC), Lys (AAG), Trp (TGG), Ser (TCC), Arg (AGG), Val (GTG), Pro (CCC), Thr (ACC), Glu (GAG); Leu (CTG), His (CAC), Ile (ATC), Asn (AAC), Cys (TGC), Ala (GCC), Gln (CAG), Phe (TTC) and Tyr (TAC). For an additional discussion relating to mammalian (human) codon optimization, see WO 97/31115 (PCT/US97/02294), which is hereby incorporated by reference. It is intended that the skilled artisan may use alternative versions of codon optimization or 10 may omit this step when generating HIV pol vaccine constructs within the scope of the present invention. Therefore, the present invention also relates to non-codon optimized versions of DNA molecules and associated DNA vaccines which encode the various wild type and modified forms of the HIV Pol protein disclosed herein. However, codon optimization of these constructs is a preferred embodiment of this 15 invention.

A particular embodiment of the present invention relates to codon optimized wt-pol DNA constructs (herein, "wt-pol" or "wt-pol (codon optimized)") wherein DNA sequences encoding the protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase H activity) and IN integrase activity. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1, the open reading frame being contained from an initiating Met residue at nucleotides 10-12 to a termination codon from nucleotides 2560-2562. SEQ ID NO:1 is as follows:

AGATCTACCA TGGCCCCCAT CTCCCCCATT GAGACTGTGC CTGTGAAGCT GAAGCCTGGC
25 ATGGATGGCC CCAAGGTGAA GCAGTGGCCC CTGACTGAGG AGAAGATCAA GGCCCTGGTG
GAAATCTGCA CTGAGATGGA GAAGGAGGGC AAAATCTCCA AGATTGGCCC CGAGAACCCC
TACAACACCC CTGTGTTTGC CATCAAGAAG AAGGACTCCA CCAAGTGGAG GAAGCTGGTG
GACTTCAGGG AGCTGAACAA GAGGACCCAG GACTTCTGGG AGGTGCAGCT GGGCATCCCC
CACCCCGCTG GCCTGAAGAA GAAGAAGTCT GTGACTGTGC TGATGTGGG GGATGCCTAC
30 TTCTCTGTGC CCCTGGATGA GGACTTCAGG AAGTACACTG CCTTCACCAT CCCCTCCATC
AACAAATGAGA CCCCTGGCAT CAGGTACCAAG TACAATGTGC TGCCCCAGGG CTGGAAAGGGC
TCCCCCTGCCA TCTTCCAGTC CTCCATGACC AAGATCCTGG AGCCCTTCAG GAAGCAGAAC
CCTGACATTG TGATCTACCA GTACATGGAT GACCTGTATG TGGGCTCTGA CCTGGAGATT
GGGCAGCACA GGACCAAGAT TGAGGAGCTG AGGCAGCACC TGCTGAGGTG GGGCCTGACC

ACCCCTGACA AGAAGCACCA GAAGGGAGCCC CCCTTCCTGT GGATGGGCTA TGAGCTGCAC
CCCGACAACTT GGACTGTGCA GCCCATTGTG CTGCCTGAGA AGGACTCCTG GACTGTGAAT
GACATCCAGA AGCTGGTGGG CAAGCTGAAC TGGGCCTCCC AAATCTACCC TGGCATCAAG
GTGAGGCAGC TGTGCAAGCT GCTGAGGGGC ACCAAGGCC TGACTGAGGT GATCCCCCTG
5 ACTGAGGAGG CTGAGCTGGA GCTGGCTGAG AACAGGGAGA TCCTGAAGGA GCCTGTGCAT
GGGGTGTACT ATGACCCCTC CAAGGACCTG ATTGCTGAGA TCCAGAAGCA GGGCCAGGGC
CAGTGGACCT ACCAAATCTA CCAGGAGCCC TTCAAGAACCC TGAAGACTGG CAAGTATGCC
AGGATGAGGG GGGCCCACAC CAATGATGTG AAGCAGCTGA CTGAGGCTGT GCAGAAGATC
ACCACTGAGT CCATTGTGAT CTGGGGCAAG ACCCCCAAGT TCAAGCTGCC CATCCAGAAG
10 GAGACCTGGG AGACCTGGTG GACTGAGTAC TGGCAGGCCA CCTGGATCCC TGAGTGGGAG
TTTGTGAACA CCCCCCCCCCT GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATTGTG
GGGGCTGAGA CCTTCTATGT GGATGGGCT GCCAACAGGG AGACCAAGCT GGGCAAGGCT
GGCTATGTGA CCAACAGGGG CAGGCAGAAG GTGGTGACCC TGACTGACAC CACCAACCAG
. AAGACTGAGC TCCAGGCCAT CTACCTGGCC CTCCAGGACT CTGGCCTGGA GGTGAACATT
15 GTGACTGACT CCCAGTATGC CCTGGGCATC ATCCAGGCC ACCCTGATCA GTCTGAGTCT
GAGCTGGTGA ACCAGATCAT TGAGCAGCTG ATCAAGAACCC AGAAGGTGTA CCTGGCCTGG
GTGCCTGCC ACAAGGGCAT TGGGGCAAT GAGCAGGTGG ACAAGCTGGT GTCTGCTGGC
ATCAGGAAGG TGCTGTTCCCT GGATGGCATT GACAAGGCC AGGATGAGCA TGAGAAGTAC
CACTCCAACCT GGAGGGCTAT GGCTCTGAC TTCAACCTGC CCCCTGTGGT GGCTAAGGAG
20 ATTGTGGCCT CCTGTGACAA GTGCCAGCTG AAGGGGGAGG CCATGCATGG GCAGGGTGGAC
TGCTCCCCCTG GCATCTGGCA GCTGGACTGC ACCCACCTGG AGGGCAAGGT GATCCTGGTG
GCTGTGCATG TGGCCTCCGG CTACATTGAG GCTGAGGTGA TCCCTGCTGA GACAGGCCAG
GAGACTGCCT ACTTCCTGCT GAAGCTGGCT GGCAGGTGGC CTGTGAAGAC CATCCACACT
GACAATGGCT CCAACTTCAC TGGGCCACA GTGAGGGCTG CCTGCTGGTG GGCTGGCATC
25 AAGCAGGAGT TTGGCATCCC CTACAACCCC CAGTCCCAGG GGGTGGTGGA GTCCATGAAC
AAGGAGCTGA AGAAGATCAT TGGGCAGGTG AGGGACCAGG CTGAGCACCT GAAGACAGCT
GTGCAGATGG CTGTGTTCAT CCACAACCTTC AAGAGGAAGG GGGGCATCGG GGGCTACTCC
GCTGGGGAGA GGATTGTGGA CATCATTGCC ACAGACATCC AGACCAAGGA GCTCCAGAAG
CAGATCACCA AGATCCAGAA CTTCAAGGGTG TACTACAGGG ACTCCAGGAA CCCCTGTGG
30 AAGGGCCCTG CCAAGCTGCT GTGGAAGGGG GAGGGGGCTG TGGTGTGATCCA GGACAACCT
GACATCAAGG TGGTGCCAG GAGGAAGGCC AAGATCATCA GGGACTATGG CAAGCAGATG
GCTGGGGATG ACTGTGTGGC CTCCAGGCAG GATGAGGACT AAAGCCCCGGG CAGATCT (SEQ
ID NO:1).

The open reading frame of the wild type pol construct disclosed as SEQ ID NO:1 contains 850 amino acids, disclosed herein as SEQ ID NO:2, as follows:

Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
5 Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala
Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Asp
10 Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly
15 Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln
Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
20 Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr
Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
25 Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
Glu Thr Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
30 Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala
Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu
Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro

Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys
Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
5 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
10 Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu
Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
15 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
20 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
Glu Asp (SEQ ID NO:2).

The present invention especially relates to a codon optimized HIV-1 DNA pol construct wherein, in addition to deletion of the portion of the wild type sequence encoding the protease activity, a combination of active site residue mutations are introduced which are deleterious to HIV-1 pol (RT-RH-IN) activity of the expressed protein. Therefore, the present invention preferably relates to a HIV-1 DNA pol construct which is devoid of DNA sequences encoding any PR activity, as well as containing a mutation(s) which at least partially, and preferably substantially, abolishes RT, RNase and/or IN activity. One type of HIV-1 pol mutant may include but is not limited to a mutated DNA molecule comprising at least one nucleotide substitution which results in a point mutation which effectively alters an active site within the RT, RNase and/or IN regions of the expressed protein, resulting in at least substantially decreased enzymatic activity for the RT, RNase H and/or IN functions of

HIV-1 Pol. In a preferred embodiment of this portion of the invention, a HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct in a DNA molecule which contains at least one point mutation

5 which alters the active site of the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. To this end, an especially preferred HIV-1 DNA pol construct is exemplified herein and contains nine codon substitution

10 mutations which results in an inactivated Pol protein (IA Pol: SEQ ID NO:4, Figure 2A-C) which has no PR, RT, RNase or IN activity, wherein three such point mutations reside within each of the RT, RNase and IN catalytic domains. Therefore, an especially preferred exemplification is a DNA molecule which encodes IA-pol, which contains all nine mutations as shown below in Table 1. An additional preferred

15 amino acid residue for substitution is Asp551, localized within the RNase domain of Pol. Any combination of the mutations disclosed herein may suitable and therefore may be utilized as an IA-Pol-based vaccine of the present invention. While addition and deletion mutations are contemplated and within the scope of the invention, the preferred mutation is a point mutation resulting in a substitution of the wild type

20 amino acid with an alternative amino acid residue.

Table 1

	<u>wt aa</u>	<u>aa residue</u>	<u>mutant aa</u>	<u>enzyme function</u>
25	Asp	112	Ala	RT
	Asp	187	Ala	RT
	Asp	188	Ala	RT
	Asp	445	Ala	RNase H
30	Glu	480	Ala	RNase H
	Asp	500	Ala	RNase H
	Asp	626	Ala	IN
	Asp	678	Ala	IN
	Glu	714	Ala	IN

It is preferred that point mutations be incorporated into the IApol mutant vaccines of the present invention so as to lessen the possibility of altering epitopes in and around the active site(s) of HIV-1 Pol.

To this end, SEQ ID NO:3 discloses the nucleotide sequence which codes for 5 a codon optimized pol in addition to the nine mutations shown in Table 1, disclosed as follows, and referred to herein as "IApol":

```

AGATCTACCA TGGCCCCCAT CTCCCCCATT GAGACTGTGC CTGTGAAGCT GAAGCCTGGC
ATGGATGGCC CCAAGGTGAA GCAGTGGCCC CTGACTGAGG AGAAGATCAA GGCCCTGGTG
GAAATCTGCA CTGAGATGGA GAAGGAGGGC AAAATCTCCA AGATTGGCCC CGAGAACCCC
10 TACAACACCC CTGTGTTGC CATCAAGAAG AAGGACTCCA CCAAGTGGAG GAAGCTGGTG
GACTTCAGGG AGCTGAACAA GAGGACCCAG GACTTCTGGG AGGTGCAGCT GGGCATCCCC
CACCCCGCTG GCCTGAAGAA GAAGAAGTCT GTGACTGTGC TGGCTGTGGG GGATGCCTAC
TTCTCTGTGC CCCTGGATGA GGACTTCAGG AAGTACACTG CCTTCACCAT CCCCTCCATC
AACAAATGAGA CCCCTGGCAT CAGGTACCAAG TACAATGTGC TGCCCCAGGG CTGGAAGGGC
15 TCCCCCTGCCA TCTTCCAGTC CTCCATGACC AAGATCCTGG AGCCCTTCAG GAAGCAGAAC
CCTGACATTG TGATCTACCA GTACATGGCT GCCCTGTATG TGGGCTCTGA CCTGGAGATT
GGGCAGCACCA GGACCAAGAT TGAGGAGCTG AGGCAGCACC TGCTGAGGTG GGGCCTGACC
ACCCCTGACA AGAAGCACCA GAAGGAGCCC CCCTTCCTGT GGATGGGCTA TGAGCTGCAC
CCCGACAAGT GGACTGTGCA GCCCATTGTG CTGCCTGAGA AGGACTCCCTG GACTGTGAAT
20 GACATCCAGA AGCTGGGGGG CAAGCTGAAC TGGGCCTCCC AAATCTACCC TGGCATCAAG
GTGAGGCAGC TGTGCAAGCT GCTGAGGGGC ACCAAGGGCC TGACTGAGGT GATCCCCCTG
ACTGAGGAGG CTGAGCTGGA GCTGGCTGAG AACAGGGAGA TCCTGAAGGA GCCTGTGCAT
GGGGTGTACT ATGACCCCTC CAAGGACCTG ATTGCTGAGA TCCAGAAGCA GGGCCAGGGC
CACTGGACCT ACCAAATCTA CCAGGAGCCC TTCAAGAACCC TGAAAGACTGG CAACTATGCC
25 AGGATGAGGG GGGCCCACAC CAATGATGTG AAGCAGCTGA CTGAGGCTGT GCAGAAGATC
ACCACTGAGT CCATTGTGAT CTGGGGCAAG ACCCCCAAGT TCAAGCTGCC CATCCAGAAC
GAGACCTGGG AGACCTGGTG GACTGAGTAC TGGCAGGCCA CCTGGATCCC TGAGTGGGAG
TTTGTGAACA CCCCCCCCCCT GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATTGTG
GGGGCTGAGA CCTTCTATGT GGCTGGGGCT GCCAACAGGG AGACCAAGCT GGGCAAGGCT
30 GGCTATGTGA CCAACAGGGG CAGGCAGAACAG GTGGTGACCC TGACTGACAC CACCAACCAG
AAGACTGCC TCCAGGCCAT CTACCTGGCC CTCCAGGACT CTGGCCTGGA GGTGAACATT
GTGACTGCCT CCCAGTATGC CCTGGGCATC ATCCAGGGCC AGCCTGATCA GTCTGAGTCT
GAGCTGGTGA ACCAGATCAT TGAGCAGCTG ATCAAGAAGG AGAAGGTGTA CCTGGCCTGG
GTGCCTGCCA ACAAGGGCAT TGGGGCAAT GAGCAGGTGG ACAAGCTGGT GTCTGCTGGC

```

ATCAGGAAGG TGCTGTTCCT GGATGGCATT GACAAGGCC AGGATGAGCA TGAGAACTAC
CACTCCAATC GGAGGGCTAT GGCCCTCTGAC TTCAACCTGC CCCCTGTGGT GGCTAAGGAG
ATTGTGGCCT CCTGTGACAA GTGCCAGCTG AAGGGGGAGG CCATGCATGG CCAGGTGGAC
TGCTCCCCCTG GCATCTGGCA GCTGGCCTGC ACCCACCTGG AGGGCAAGGT GATCCTGGTG
5 GCTGTGCATG TGGCCTCCGG CTACATTGAG GCTGAGGTGA TCCCTGCTGA GACAGGCCAG
GAGACTGCCT ACTTCCTGCT GAAGCTGGCT GGCAGGTGGC CTGTGAAGAC CATCCACACT
GCCAATGGCT CCAACTTCAC TGGGGCCACA GTGAGGGCTG CCTGCTGGTG GGCTGGCATC
AAGCAGGAGT TTGGCATCCC CTACAACCCC CAGTCCCAGG GGGTGGTGGC CTCCATGAAC
AAGGAGCTGA AGAAGATCAT TGGGCAGGTG AGGGACCAGG CTGAGCACCT GAAGACAGCT
10 GTGCAGATGG CTGTGTTCAT CCACAACCTTC AAGAGGAAGG GGGGCATCGG GGGCTACTCC
GCTGGGGAGA GGATTGTGGA CATCATTGCC ACAGACATCC AGACCAAGGA GCTCCAGAAG
CAGATCACCA AGATCCAGAA CTTCAGGGTG TACTACAGGG ACTCCAGGAA CCCCCTGTGG
AAGGGCCCTG CCAAGCTGCT GTGGAAGGGG GAGGGGGCTG TGGTGTATCCA GGACAACCT
GACATCAAGG TGGTGCCCAG GAGGAAGGCC AAGATCATCA GGGACTATGG CAAGCAGATG
15 GCTGGGGATG ACTGTGTGGC CTCCAGGCAG GATGAGGACT AAAGCCCGGG CAGATCT (SEQ ID
NO:3).

In order to produce the IA-pol DNA vaccine construction, inactivation of the enzymatic functions was achieved by replacing a total of nine active-site residues from the enzyme subunits with alanine side-chains. As shown in Table 1, all residues
20 that comprise the catalytic triad of the polymerase, namely Asp112, Asp187, and Asp188, were substituted with alanine (Ala) residues (Larder, et al., *Nature* 1987, 327: 716-717; Larder, et al., 1989, *Proc. Natl. Acad. Sci.* 1989, 86: 4803-4807). Three additional mutations were introduced at Asp445, Glu480 and Asp500 to abolish RNase H activity (Asp551 was left unchanged in this IA Pol construct), with each
25 residue being substituted for an Ala residue, respectively (Davies, et al., 1991, *Science* 252: 88-95; Schatz, et al., 1989, *FEBS Lett.* 257: 311-314; Mizrahi, et al., 1990, *Nucl. Acids. Res.* 18: pp. 5359-5353). HIV pol integrase function was abolished through three mutations at Asp626, Asp678 and Glu714. Again, each of these residues has been substituted with an Ala residue (Wiskerchen, et al., 1995, *J.
30 Virol.* 69: 376-386; Leavitt, et al., 1993, *J. Biol. Chem.* 268: 2113-2119). Amino acid residue Pro3 of SEQ ID NO:4 marks the start of the RT gene. The complete amino acid sequence of IA-Pol is disclosed herein as SEQ ID NO:4, as follows:
Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys

Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala
Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
5 Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Ala
Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
10 Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly
Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln
Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
15 Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr
Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
20 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
Glu Thr Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
25 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala
Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala
Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
30 Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys
Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys

Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
5 Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu
10 Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
15 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
Glu Asp (SEQ ID NO:4).

As noted above, it will be understood that any combination of the mutations disclosed above may be suitable and therefore be utilized as an IA-pol-based vaccine of the present invention. For example, it may be possible to mutate only 2 of the 3 residues within the respective reverse transcriptase, RNase H, and integrase coding regions while still abolishing these enzymatic activities. However, the IA-pol construct described above and disclosed as SEQ ID NO:3, as well as the expressed protein (SEQ ID NO:4) is preferred. It is also preferred that at least one mutation be present in each of the three catalytic domains.

Another aspect of the present invention is to generate codon optimized HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as from tPA (tissue-type plasminogen activator) or by a leader peptide such as is found in highly expressed mammalian proteins such as immunoglobulin leader peptides. Any functional leader peptide may be tested for efficacy. However, a preferred embodiment of the present invention is to provide for HIV-1 Pol mutant vaccine constructions as disclosed herein which also comprise a leader peptide, preferably a leader peptide from human tPA. In other words, a codon optimized

HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) may also comprise a leader peptide at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. As shown in Figure 1A-B for the DNA vector V1Jns, a DNA vector which may be utilized to practice the present invention may be modified by known recombinant DNA methodology to contain a leader signal peptide of interest, such that downstream cloning of the modified HIV-1 protein of interest results in a nucleotide sequence which encodes a modified HIV-1 tPA/Pol protein. In the alternative, as noted above, insertion of a nucleotide sequence which encodes a leader peptide may be inserted into a DNA vector housing the open reading frame for the Pol protein of interest. Regardless of the cloning strategy, the end result is a polynucleotide vaccine which comprises vector components for effective gene expression in conjunction with nucleotide sequences which encode a modified HIV-1 Pol protein of interest, including but not limited to a HIV-1 Pol protein which contains a leader peptide. The amino acid sequence of the human tPA leader utilized herein is as follows:

MDAMKRGLCCVLLLCGAVFVSPSEISS (SEQ ID NO:28). Therefore, another aspect of the present invention is to generate HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as from tPA. To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame disclosed herein as SEQ ID NO:6.

To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region (herein, "tPA-wt-pol"). A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame being contained from an initiating Met residue at nucleotides 8-10 to a termination codon from nucleotides 2633-2635. SEQ ID NO:5 is as follows:

GATCACCATG GATGCAATGA AGAGAGGGCT CTGCTGTGTG CTGCTGCTGT GTGGAGCAGT
CTTCGTTTCG CCCAGCGAGA TCTCCGCCCT CATCTCCCC ATTGAGACTG TGCCCTGTGAA
GCTGAAGCCT GGCATGGATG GCCCCAAGGT GAAGCAGTGG CCCCTGACTG AGGAGAAGAT
CAAGGCCCTG GTGGAAATCT GCACTGAGAT GGAGAAGGAG GGCAAATCT CCAAGATTGG

CCCCGAGAAC CCCTACAACA CCCCTGTGTT TGCCATCAAG AAGAAGGACT CCACCAAGTG
GAGGAAGCTG GTGGACTTCA GGGAGCTGAA CAAGAGGACC CAGGACTTCT GGGAGGTGCA
GCTGGGCATC CCCCACCCCG CTGGCCTGAA GAAGAAGAAG TCTGTGACTG TGCTGGATGT
GGGGGATGCC TACTTCTCTG TGCCCCTGGA TGAGGACTTC AGGAAGTACA CTGCCTTCAC
5 CATCCCCCTCC ATCAACAATG AGACCCCTGG CATCAGGTAC CAGTACAATG TGCTGCCCA
GGGCTGGAAG GGCTCCCCCTG CCATCTTCCA GTCCTCCATG ACCAAGATCC TGGAGCCCTT
CAGGAAGCAG AACCCCTGACA TTGTGATCTA CCAGTACATG GATGACCTGT ATGTGGGCTC
TGACCTGGAG ATTGGGCAGC ACAGGACCAA GATTGAGGAG CTGAGGCAGC ACCTGCTGAG
GTGGGGCCTG ACCACCCCTG ACAAGAAGCA CCAGAAGGAG CCCCCCTTCC TGTGGATGGG
10 CTATGAGCTG CACCCCGACA AGTGGACTGT GCAGCCCCATT GTGCTGCCCTG AGAAGGACTC
CTGGACTGTG AATGACATCC AGAAGCTGGT GGGCAAGCTG AACTGGGCCT CCCAAATCTA
CCCTGGCATC AAGGTGAGGC AGCTGTGCAA GCTGCTGAGG GGCACCAAGG CCCTGACTGA
GGTGATCCCC CTGACTGAGG AGGCTGAGCT GGAGCTGGCT GAGAACAGGG AGATCCTGAA
GGAGCCTGTG CATGGGGTGT ACTATGACCC CTCCAAGGAC CTGATTGCTG AGATCCAGAA
15 GCAGGGCCAG GGCCAGTGGAA CCTACCAAAT CTACCAGGAG CCCTCAAGA ACCTGAAGAC
TGGCAAGTAT GCCAGGATGA GGGGGGCCCA CACCAATGAT GTGAAGCAGC TGACTGAGGC
TGTGCAGAAG ATCACCACTG AGTCCATTGT GATCTGGGGC AAGACCCCCA AGTTCAAGCT
GCCCATCCAG AAGGAGACCT GGGAGACCTG GTGGACTGAG TACTGGCAGG CCACCTGGAT
CCCTGAGTGG GAGTTGTGA ACACCCCCCCC CCTGGTGAAG CTGTGGTACC AGCTGGAGAA
20 GGAGCCCCATT GTGGGGCCTG AGACCTTCTA TGTGGATGGG GCTGCCAACAGGGAGACCAA
GCTGGCAAG GCTGGCTATG TGACCAACAG GGGCAGGCAG AAGGTGGTGA CCCTGACTGA
CACCAACAC CAGAAGACTG AGCTCCAGGC CATCTACCTG GCCCTCCAGG ACTCTGGCCT
GGAGGTGAAC ATTGTGACTG ACTCCCAGTA TGCCCTGGGC ATCATCCAGG CCCAGCCTGA
TCAGTCTGAG TCTGAGCTGG TGAACCAAGAT CATTGAGCAG CTGATCAAGA AGGAGAAGGT
25 GTACCTGGCC TGGGTGCCCTG CCCACAAGGG CATTGGGGC AATGAGCAGG TGGACAAGCT
GGTGTCTGCT GGCATCAGGA AGGTGCTGTT CCTGGATGGC ATTGACAAGG CCCAGGATGA
GCATGAGAAG TACCACTCCA ACTGGAGGGC TATGGCCTCT GACTTCACCC TGCCCCCTGT
GGTGGCTAACAG GAGATTGTGG CCTCCTGTGA CAAGTGCCAG CTGAAGGGGG AGGCCATGCA
TGGGAGGTG GACTGCTCCC CTGGCATCTG GCAGCTGGAC TGCACCCACC TGGAGGGCAA
30 GGTGATCCTG GTGGCTGTGC ATGTGGCCTC CGGCTACATT GAGGCTGAGG TGATCCCTGC
TGAGACAGGC CAGGAGACTG CCTACTTCCT GCTGAAGCTG GCTGGCAGGT GGCCTGTGAA
GACCATCCAC ACTGACAATG GCTCCAACCTT CACTGGGCC ACAGTGAGGG CTGCCTGCTG
GTGGGCTGGC ATCAAGCAGG AGTTGGCAT CCCCTACAAC CCCCAGTCCC AGGGGGTGGT
GGAGTCCATG AACAAAGGAGC TGAAGAAGAT CATTGGCAG GTGAGGGACC AGGCTGAGCA

CCTGAAGACA GCTGTGCAGA TGGCTGTGTT CATCCACAAAC TTCAAGAGGA AGGGGGGCAT
CGGGGGCTAC TCCGCTGGGG AGAGGATTGT GGACATCATT GCCACAGACA TCCAGACCAA
GGAGCTCCAG AAGCAGATCA CCAAGATCCA GAACTTCAGG GTGTACTACA GGGACTCCAG
GAACCCCCTG TGGAAGGGCC CTGCCAAGCT GCTGTGGAAG GGGGAGGGGG CTGTGGTGAT
5 CCAGGACAAC TCTGACATCA AGGTGGTGC CAGGAGGAAG GCCAAGATCA TCAGGGACTA
TGGCAAGCAG ATGGCTGGGG ATGACTGTGT GGCCCTCCAGG CAGGATGAGG ACTAAAGCCC
GGGCAGATCT (SEQ ID NO:5).

The open reading frame of the wild type tPA-pol construct disclosed as SEQ ID NO:5 contains 875 amino acids, disclosed herein as SEQ ID NO:6, as follows:

10 Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
15 Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser
Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
20 Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
25 Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
30 Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
Ile Ala Glu Ile Gln Lys Gln Gly Gln Gln Trp Thr Tyr Gln Ile
Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln

Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr
Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
5 Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala
Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr
Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
10 Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
15 Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu
Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn
20 Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala
Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
Val Val Glu Ser Met Asn Lys Glu Leu Lys Ile Ile Gly Gln Val
Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly
25 Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
30 Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:6).

The present invention also relates to a codon optimized HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) which comprises a leader peptide at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. Any such HIV-1 DNA

pol mutant disclosed in the above paragraphs is suitable for fusion downstream of a leader peptide, such as a leader peptide including but not limited to the human tPA leader sequence. Therefore, any such leader peptide-based HIV-1 pol mutant construct may include but is not limited to a mutated DNA molecule which effectively

5 alters the catalytic activity of the RT, RNase and/or IN region of the expressed protein, resulting in at least substantially decreased enzymatic activity one or more of the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a leader peptide/HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and

10 IN activity. An especially preferable HIV-1 DNA pol construct is a DNA molecule which contains at least one point mutation which alters the active site and catalytic activity within the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished, and preferably totally abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each

15 catalytic domain responsible for RT, RNase H and IN activity, respectfully. An especially preferred embodiment of this portion of the invention relates to a human tPA leader fused to the IA-Pol protein comprising the nine mutations shown in Table 1. The DNA molecule is disclosed herein as SEQ ID NO:7 and the expressed tPA-IA Pol protein comprises a fusion junction as shown in Figure 3. The complete amino acid sequence of the expressed protein is set forth in SEQ ID NO:8. To this end, SEQ 20 ID NO:7 discloses the nucleotide sequence which codes for a human tPA leader fused to the IA Pol protein comprising the nine mutations shown in Table 1 (herein, "tPA-opt-IAPol"). The open reading frame begins with the initiating Met (nucleotides 8-10) and terminates with a "TAA" codon at nucleotides 2633-2635. The nucleotide 25 sequence encoding tPA-IAPol is also disclosed as follows:

GATCACCATG GATGCAATGA AGAGAGGGCT CTGCTGTG TGCTGCTGT GTGGAGCAGT
CTTCGTTTCG CCCAGCGAGA TCTCCGCCCC CATCTCCCCC ATTGAGACTG TGCCCTGTGAA
GCTGAAGCCT GGCATGGATG GCCCCAAGGT GAAGCAGTGG CCCCTGACTG AGGAGAAGAT
CAAGGCCCTG GTGGAAATCT GCACTGAGAT GGAGAAGGAG GGCAAATCT CCAAGATTGG
30 CCCCGAGAAC CCCTACAACA CCCCTGTGTT TGCCATCAAG AAGAAGGACT CCACCAAGTG
GAGGAAGCTG GTGGACTTCA GGGAGCTGAA CAAGAGGACC CAGGACTTCT GGGAGGTGCA
GCTGGGCATC CCCCACCCCG CTGGCCTGAA GAAGAAGAAG TCTGTGACTG TGCTGGCTGT
GGGGGATGCC TACTTCCTG TGCCCCCTGGA TGAGGACTTC AGGAAGTACA CTGCCTTCAC
CATCCCCCTCC ATCAACAATG AGACCCCTGG CATCAGGTAC CAGTACAATG TGCTGCCCA

GGGCTGGAAG GGCTCCCTG CCATCTCCA GTCCTCCATG ACCAAGATCC TGGAGCCCT
 CAGGAAGCAG AACCCCTGACA TTGTGATCTA CCAGTACATG GCTGCCCTGT ATGTGGGCTC
 TGACCTGGAG ATTGGGCAGC ACAGGACCAA GATTGAGGAG CTGAGGCAGC ACCTGCTGAG
 GTGGGGCTG ACCACCCCTG ACAAGAACCA CCAGAAGGAG CCCCTTCC TGTGGATGGG
 5 CTATGAGCTG CACCCCGACA AGTGGACTGT GCAGCCCATT GTGCTGCCTG AGAAGGACTC
 CTGGACTGTG AATGACATCC AGAAGCTGGT GGGCAAGCTG AACTGGCCT CCCAAATCTA
 CCCTGGCATC AAGGTGAGGC AGCTGTGCAA GCTGCTGAGG GGCACCAAGG CCCTGACTGA
 GGTGATCCCC CTGACTGAGG AGGCTGAGCT GGAGCTGGCT GAGAACAGGG AGATCCTGAA
 GGAGCCTGTG CATGGGGTGT ACTATGACCC CTCCAAGGAC CTGATTGCTG AGATCCAGAA
 10 GCAGGGCCAG GGCCAGTGA CCTACCAAAT CTACCAGGAG CCCTCAAGA ACCTGAAGAC
 TGGCAAGTAT GCCAGGATGA GGGGGGCCA CACCAATGAT GTGAAGCAGC TGACTGAGGC
 TGTGCAGAAG ATCACCACTG AGTCCATTGT GATCTGGGC AAGACCCCCA AGTTCAAGCT
 GCCCATCCAG AAGGAGACCT GGGAGACCTG GTGGACTGAG TACTGGCAGG CCACCTGGAT
 CCCTGAGTGG GAGTTTGTGA ACACCCCCCC CCTGGTGAAG CTGTGGTACC AGCTGGAGAA
 15 GGAGCCCATT GTGGGGCTG AGACCTCTA TGTGGCTGGG GCTGCCAACA GGGAGACCAA
 GCTGGCAAG GCTGGCTATG TGACCAACAG GGGCAGGCAG AAGGTGGTGA CCCTGACTGA
 CACCACCAAC CAGAAGACTG CCCTCCAGGC CATCTACCTG GCCCTCCAGG ACTCTGGCCT
 GGAGGTGAAC ATTGTGACTG CCTCCCAGTA TGCCCTGGGC ATCATCCAGG CCCAGCCTGA
 TCAGTCTGAG TCTGAGCTGG TGAACCAGAT CATTGAGCAG CTGATCAAGA AGGAGAAGGT
 20 GTACCTGGCC TGGGTGCCTG CCCACAAGGG CATTGGGGC AATGAGCAGG TGGACAAGCT
 GGTGTCTGCT GGCATCAGGA AGGTGCTGTT CCTGGATGGC ATTGACAAGG CCCAGGATGA
 GCATGAGAAG TACCACTCCA ACTGGAGGGC TATGGCCTCT GACTTCAACC TGCCCCCTGT
 GGTGGCTAAG GAGATTGTGG CCTCCTGTGA CAAGTGCCAG CTGAAGGGGG AGGCCATGCA
 TGGCAGGTG GACTGCTCCC CTGGCATCTG CGAGCTGGCC TGCACCCACC TGGAGGGCAA
 25 GGTGATCCTG GTGGCTGTGC ATGTGGCCTC CGGCTACATT GAGGCTGAGG TGATCCCTGC
 TGAGACAGGC CAGGAGACTG CCTACTTCCT GCTGAAGCTG GCTGGCAGGT GCCCTGTGAA
 GACCATCCAC ACTGCCAATG GCTCCAACCT CACTGGGGC ACAGTGAGGG CTGCCTGCTG
 GTGGGCTGGC ATCAAGCAGG AGTTGGCAT CCCCTACAAC CCCCAGTCCC AGGGGGTGGT
 GCCCTCCATG AACAAAGGAGC TGAAGAAGAT CATTGGCGAG GTGAGGGACC AGGCTGAGCA
 30 CCTGAAGACA GCTGTGCAGA TGGCTGTGTT CATCCACAAAC TTCAAGAGGA AGGGGGGCAT
 CGGGGGCTAC TCCGCTGGGG AGAGGATTGT GGACATCATT GCCACAGACA TCCAGACCAA
 GGAGCTCCAG AAGCAGATCA CCAAGATCCA GAACTTCAGG GTGTACTACA GGGACTCCAG
 GAACCCCCCTG TGGAAAGGGCC CTGCCAAGCT GCTGTGGAAG GGGGAGGGGG CTGTGGTGT
 CCAGGACAAC TCTGACATCA AGGTGGTGCC CAGGAGGAAG GCCAAGATCA TCAGGGACTA

TGGCAAGCAG ATGGCTGGGG ATGACTGTGT GGCCTCCAGG CAGGATGAGG ACTAAAGCCC
GGGCAGATCT (SEQ ID NO:7).

The open reading frame of the tPA-IA-pol construct disclosed as SEQ ID NO:7 contains 875 amino acids, disclosed herein as tPA-IA-Pol and SEQ ID NO:8, as follows:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
10 Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser
15 Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
20 His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
25 Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
Ile Ala Glu Ile Gln Lys Gln Gly Gln Gln Trp Thr Tyr Gln Ile
Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
30 Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln
Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr
Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala

Glu Thr Phe Tyr Val Ala Ala Asn Arg Glu Thr Lys Leu Gly
Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala
Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr
5 Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
10 Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu
Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
15 Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn
Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala
Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
Val Val Ala Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val
Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
20 Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly
Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
25 Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:8).

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified proteins set forth below as SEQ ID NOS: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors are well known in the art and the present DNA vector vaccines may be comprised of any such expression backbone which contains at least a promoter for RNA polymerase

transcription, and a transcriptional terminator 3' to the HIV pol coding sequence. In one preferred embodiment, the promoter is the Rous sarcoma virus (RSV) long terminal repeat (LTR) which is a strong transcriptional promoter. A more preferred promoter is the cytomegalovirus promoter with the intron A sequence (CMV-intA).

5 A preferred transcriptional terminator is the bovine growth hormone terminator. In addition, to assist in large scale preparation of an HIV pol DNA vector vaccine, an antibiotic resistance marker is also preferably included in the expression vector. Ampicillin resistance genes, neomycin resistance genes or any other pharmaceutically acceptable antibiotic resistance marker may be used. In a preferred embodiment of

10 this invention, the antibiotic resistance gene encodes a gene product for neomycin resistance. Further, to aid in the high level production of the pharmaceutical by fermentation in prokaryotic organisms, it is advantageous for the vector to contain an origin of replication and be of high copy number. Any of a number of commercially available prokaryotic cloning vectors provide these benefits. In a preferred

15 embodiment of this invention, these functionalities are provided by the commercially available vectors known as pUC. It is desirable to remove non-essential DNA sequences. Thus, the lacZ and lacI coding sequences of pUC are removed in one embodiment of the invention.

DNA expression vectors which exemplify but in no way limit the present invention are disclosed in PCT International Application No. PCT/US94/02751, International Publication No. WO 94/21797, hereby incorporated by reference. A first DNA expression vector is the expression vector pnRSV, wherein the rous sarcoma virus (RSV) long terminal repeat (LTR) is used as the promoter. A second embodiment relates to plasmid V1, a mutated pBR322 vector into which the CMV promoter and the BGH transcriptional terminator is cloned. Another embodiment regarding DNA vector backbones relates to plasmid V1J. Plasmid V1J is derived from plasmid V1 and removes promoter and transcription termination elements in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. Therefore, V1J also contains the CMVintA promoter and (BGH) transcription termination elements which control the expression of the HIV pol-based genes disclosed herein. The backbone of V1J is provided by pUC18. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. The entire lac operon was removed and the remaining plasmid was purified from an agarose electrophoresis gel,

blunt-ended with the T4 DNA polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element. In a preferred DNA expression vector, the ampicillin resistance gene is removed from V1J and replaced with a neomycin resistance gene, to generate V1Jneo. An especially preferred DNA 5 expression vector is V1Jns, which is the same as V1J except that a unique Sfi1 restriction site has been engineered into the single Kpn1 site at position 2114 of V1Jneo. The incidence of Sfi1 sites in human genomic DNA is very low (approximately 1 site per 100,000 bases). Thus, this vector allows careful monitoring for expression vector integration into host DNA, simply by Sfi1 digestion of extracted genomic 10 DNA. Yet another preferred DNA expression vector used as the backbone to the HIV-1 pol-based DNA vaccines of the present invention is V1R. In this vector, as much non-essential DNA as possible is "trimmed" from the vector to produce a highly compact vector. This vector is a derivative of V1Jns. This vector allows larger inserts to be used, with less concern that undesirable sequences are encoded and 15 optimizes uptake by cells when the construct encoding specific influenza virus genes is introduced into surrounding tissue. The specific DNA vectors of the present invention include but are not limited to V1, V1J (SEQ ID NO:13), V1Jneo (SEQ ID NO:14), V1Jns (Figure 1A, SEQ ID NO:15), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, 20 preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and 25 the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding HIV-1 Pol or biologically active Pol modifications or Pol-containing fusion proteins which are ligated within an appropriate DNA plasmid vector, with or without 30 a nucleotide sequence encoding a functional leader peptide. DNA vaccines of the present invention may comprise codon optimized DNA molecules encoding HIV-1 Pol or biologically active Pol modifications or Pol-containing fusion proteins ligated in DNA vectors V1, V1J (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), or any of the aforementioned vectors.

wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28. To this end, polynucleotide vaccine constructions include , V1Jns-wtpol and V1R-wtpol

5 (comprising the DNA molecule encoding WT Pol, as set forth in SEQ ID NO:2), V1Jns-tPA-WTPol, (comprising the DNA molecule encoding tPA Pol, as set forth in SEQ ID NO:6), V1Jns-IAPol (comprising the DNA molecule encoding IA Pol, as set forth in SEQ ID NO:4), and V1Jns-tPA-IAPol, (comprising the DNA molecule encoding tPA-IA Pol, as set forth in SEQ ID NO:8). Polynucleotide vaccine

10 constructions V1R-wtpol, V1Jns-IAPol, and V1Jns-tPA-IAPol, are exemplified in Example Sections 3-5.

It will be evident upon review of the teaching within this specification that numerous vector/Pol antigen constructs may be generated. While the exemplified constructs are preferred, any number of vector/Pol antigen combinations are within

15 the scope of the present invention, especially wild type or modified/inactivated Pol proteins which comprise at least one, preferably 5 or more and especially all nine mutations as shown in Table 1, with or without the inclusion of a leader sequence such as human tPA.

The DNA vector vaccines of the present invention may be formulated in any

20 pharmaceutically effective formulation for host administration. Any such formulation may be, for example, a saline solution such as phosphate buffered saline (PBS). It will be useful to utilize pharmaceutically acceptable formulations which also provide long-term stability of the DNA vector vaccines of the present invention. During storage as a pharmaceutical entity, DNA plasmid vaccines undergo a

25 physiochemical change in which the supercoiled plasmid converts to the open circular and linear form. A variety of storage conditions (low pH, high temperature, low ionic strength) can accelerate this process. Therefore, the removal and/or chelation of trace metal ions (with succinic or malic acid, or with chelators containing multiple phosphate ligands) from the DNA plasmid solution, from the formulation buffers or

30 from the vials and closures, stabilizes the DNA plasmid from this degradation pathway during storage. In addition, inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, are useful to prevent damage of the DNA plasmid from free radical production that may still occur, even in apparently demetalated solutions. Furthermore, the buffer type, pH, salt concentration, light

exposure, as well as the type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine. Therefore, formulations that will provide the highest stability of the DNA vaccine will be one that includes a demetalated solution containing a buffer (phosphate or bicarbonate) 5 with a pH in the range of 7-8, a salt (NaCl, KCl or LiCl) in the range of 100-200 mM, a metal ion chelator (e.g., EDTA, diethylenetriaminepenta-acetic acid (DTPA), malate, inositol hexaphosphate, tripolyphosphate or polyphosphoric acid), a non-reducing free radical scavenger (e.g. ethanol, glycerol, methionine or dimethyl sulfoxide) and the highest appropriate DNA concentration in a sterile glass vial, 10 packaged to protect the highly purified, nuclease free DNA from light. A particularly preferred formulation which will enhance long term stability of the DNA vector vaccines of the present invention would comprise a Tris-HCl buffer at a pH from about 8.0 to about 9.0; ethanol or glycerol at about 3% w/v; EDTA or DTPA in a concentration range up to about 5 mM; and NaCl at a concentration from about 50 15 mM to about 500 mM. The use of such stabilized DNA vector vaccines and various alternatives to this preferred formulation range is described in detail in PCT International Application No. PCT/US97/06655 and PCT International Publication No. WO 97/40839, both of which are hereby incorporated by reference.

The DNA vector vaccines of the present invention may also be formulated 20 with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention. A number of these adjuvants are known in the art and are available for use in a DNA vaccine, including but not limited to particle bombardment using DNA-coated gold beads, co-administration of DNA vaccines with plasmid DNA expressing cytokines, chemokines, or 25 costimulatory molecules, formulation of DNA with cationic lipids or with experimental adjuvants such as saponin, monophosphoryl lipid A or other compounds which increase immunogenicity of the DNA vaccine. Another adjuvant for use in the DNA vector vaccines of the present invention are one or more forms of an aluminum phosphate-based adjuvant wherein the aluminum 30 phosphate-based adjuvant possesses a molar PO₄/Al ratio of approximately 0.9. An additional mineral-based adjuvant may be generated from one or more forms of a calcium phosphate. These mineral-based adjuvants are useful in increasing cellular and humoral responses to DNA vaccination. These mineral-based compounds for use as DNA vaccines adjuvants are disclosed in PCT International

Application No. PCT/US98/02414, PCT International Publication No. WO 98/35562, which is hereby incorporated by reference. Another preferred adjuvant is a non-ionic block copolymer which shows adjuvant activity with DNA vaccines. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. (1998, *Critical Reviews in Therapeutic Drug Carrier Systems* 15(2): 89-142) review a class of non-ionic block copolymers which show adjuvant activity. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. *id.*, disclose that certain POE-POP-POE block copolymers may be useful as adjuvants to an influenza protein-based vaccine, namely higher molecular weight POE-POP-POE block copolymers containing a central POP block having a molecular weight of over about 9000 daltons to about 20,000 daltons and flanking POE blocks which comprise up to about 20% of the total molecular weight of the copolymer (see also U.S. Reissue Patent No. 36,665, U.S. Patent No. 5,567,859, U.S. Patent No. 5,691,387, U.S. Patent No. 5,696,298 and U.S. Patent No. 5,990,241, all issued to Emanuele, et al., regarding these POE-POP-POE block copolymers). WO 96/04932 further discloses higher molecular weight POE/POP block copolymers which have surfactant characteristics and show biological efficacy as vaccine adjuvants. The above cited references within this paragraph are hereby incorporated by reference in their entirety. It is therefore within the purview of the skilled artisan to utilize available adjuvants which may increase the immune response of the polynucleotide vaccines of the present invention in comparison to administration of a non-adjuvanted polynucleotide vaccine.

The DNA vector vaccines of the present invention are administered to the host by any means known in the art, such as enteral and parenteral routes. These routes of delivery include but are not limited to intramuscular injection, intraperitoneal injection, intravenous injection, inhalation or intranasal delivery, oral delivery, sublingual administration, subcutaneous administration, transdermal administration, transcutaneous administration, percutaneous administration or any form of particle bombardment, such as a biolistic device such as a "gene gun" or by any available needle-free injection device. The preferred methods of delivery of the HIV-1 Pol-based DNA vaccines disclosed herein are intramuscular injection, subcutaneous administration and needle-free injection. An especially preferred method is

intramuscular delivery.

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1 μ g to greater than about 20 mg, and preferably in doses from about 1 mg to about 5 mg is administered directly into muscle tissue. As noted above, subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, inhalation and oral delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided in a fashion which optimizes the overall immune response to the Pol-based DNA vector vaccines of the present invention.

The aforementioned polynucleotides, when directly introduced into a vertebrate *in vivo*, express the respective HIV-1 Pol protein within the animal and in turn induce a cellular immune response within the host to the expressed Pol antigen. To this end, the present invention also relates to methods of using the HIV-1 Pol-based polynucleotide vaccines of the present invention to provide effective immunoprophylaxis, to prevent establishment of an HIV-1 infection following exposure to this virus, or as a post-HIV infection therapeutic vaccine to mitigate the acute HIV-1 infection so as to result in the establishment of a lower virus load with beneficial long term consequences. As noted above, the present invention contemplates a method of administration or use of the DNA pol-based vaccines of the present invention using any of the known routes of introducing polynucleotides into living tissue to induce expression of proteins.

Therefore, the present invention provides for methods of using a DNA pol-based vaccine utilizing the various parameters disclosed herein as well as any additional parameters known in the art, which, upon introduction into mammalian tissue induces intracellular expression of these DNA pol-based vaccines. This intracellular expression of the Pol-based immunogen induces a cellular immune response which provides a substantial level of protection against an existing HIV-1 infection or provides a substantial level of protection against a future infection in a presently uninfected host.

The following examples are provided to illustrate the present invention without, however, limiting the same hereto.

EXAMPLE 1
Vaccine Vectors

V1 – Vaccine vector V1 was constructed from pCMVIE-AKI-DHFR (Whang et al., 1987, *J. Virol.* 61: 1796). The AKI and DHFR genes were removed by cutting the vector with EcoRI and self-ligating. This vector does not contain intron A in the CMV promoter, so it was added as a PCR fragment that had a deleted internal SacI site [at 1855 as numbered in Chapman, et al., 1991, *Nuc. Acids Res.* 19: 3979]. The template used for the PCR reactions was pCMVintA-Lux, made by ligating the HindIII and NheI fragment from pCMV6a120 (see Chapman et al., *ibid.*), which includes hCMV-IE1 enhancer/promoter and intron A, into the HindIII and XbaI sites of pBL3 to generate pCMVIntBL. The 1881 base pair luciferase gene fragment (HindIII-SmaI Klenow filled-in) from RSV-Lux (de Wet et al., 1987, *Mol. Cell Biol.* 7: 725) was ligated into the SalI site of pCMVIntBL, which was Klenow filled-in and phosphatase treated. The primers that spanned intron A are: 5' primer: 5'-CTATAT AAGCAGAGCTCGTTAG-3' (SEQ ID NO:10); 3' primer: 5'-GTAGCAAA GATCTAAGGACGGTGACTGCAG-3' (SEQ ID NO:11). The primers used to remove the SacI site are: sense primer, 5'-GTATGTGTCTGAAAATGAGCG TGGAGATTGGGCTCGCAC-3' (SEQ ID NO:12) and the antisense primer, 5'-GTGCGAGCCAACTCCACGCTCATTTAGAC ACATAC-3' (SEQ ID NO:13). The PCR fragment was cut with Sac I and Bgl II and inserted into the vector which had been cut with the same enzymes.

V1J – Vaccine vector V1J was generated to remove the promoter and transcription termination elements from vector V1 in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment. pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-

characterized by sequence and function, and is of small size. The entire *lac* operon was removed from this vector by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase treated with calf intestinal alkaline phosphatase, and 5 ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in *E. coli* and was designated V1J. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher 10 expression of heterologous genes compared with V1. The nucleotide sequence of V1J is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCAGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
15 ACCATATGCC GTGTGAAATA CCCCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
20 CATACTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
TGCCCACATTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
25 CGTCAATGGG AGTTTGTGTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAAACAA
CTCCGCCCCA TTGACGCCAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
AGCTCGTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCA CCCCCCTGGC
30 TTCTTATGCA TGCTATACTG TTTTGGCTT GGGGTCTATA CACCCCCGCT TCCTCATGTT
ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC
CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTGCC ACAACTCTCT
TTATTGGCTA TATGCCAATA CACTGTCTT CAGAGACTGA CACGGACTCT GTATTTTAC
AGGATGGGCT CTCATTATT ATTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC

CCGCAGTTTT TATTAACAT AACGTGGAT CTCCACGCGA ATCTCGGTA CGTGTTCGG
 ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC
 CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACAA GTGGAGGCCA GACTTAGGCA
 CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
 5 TGAAAATGAG CTCGGGGAGC GGGCTGCAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC
 GCCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC
 CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
 GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTCCA TGGGTCTTT
 CTGCAGTCAC CGTCCTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGTTGCC
 10 CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCCCTT TCCTAATAAA
 ATGAGGAAAT TGCACTCGCAT TGTCTGAGTA GGTGTCATTTC TATTCTGGGG GGTGGGGTGG
 GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
 GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAGAAC
 AGGCACATCC CCTTCTCTGT GACACACCC GTCCACGCC CTCGTTCTTA GTTCCAGCCC
 15 CACTCATAGG ACACTCATAG CTCAGGAGGG CTCCGCCCTTC AATCCCACCC GCTAAAGTAC
 TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
 GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG
 TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
 CGCTCGGTGCG TTGGCTGCG GCGAGCGGT AACTCGTCACT CAAAGGCGGT AATACGGTTA
 20 TCCACAGAAT CAGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC
 AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCC CCCTGACGAG
 CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
 CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCCT GCCGCTTACC
 GGATACCTGT CCGCCTTCT CCCCTCGGGA AGCGTGGCCG TTTCTCAATG CTCACGCTGT
 25 AGGTATCTCA GTTCGGGTGTA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC
 GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCAA CCCGGTAAGA
 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
 GGCAGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA
 30 TCCGGCAAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTG TTTGCAAGCA GCAGATTACG
 CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTT CTACGGGGTC TGACGCTCAG
 TGGAACGAAA ACTCACGTTA AGGGATTTG GTCATGAGAT TATCAAAAG GATCTTCACC
 TAGATCCTTT TAAATTTAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT
 TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT

CGTTCATCCA TAGTTGCCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA
CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA
TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCCTGC AACTTTATCC
GCCTCCATCC AGTCTATTAA TTGGTGCCTG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT
5 AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTGGT
ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG
TGCAAAAAAG CGGTTAGCTC CTTGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA
GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA
AGATGCTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG
10 CGACCGAGTT GCTCTTGCCTT GGCCTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT
TTAAAAGTGC TCATCATTGG AAAACGTTCT TCAGGGCGAA AACTCTCAAG GATCTTACCG
CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATTTTT
ACTTTCACCA GCGTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA
ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTCAATA TTATTGAAGC
15 ATTATTCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTAA GAAAAATAAA
CAAATAGGGG TTCCGGCAC ATTTCCCCGA AAAGTGCAC CTGACGTCTA AGAAACCATT
ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA GGCCCTTTCG TC (SEQ ID
NO:14).

V1Jneo – Construction of vaccine vector V1Jneo expression vector involved
20 removal of the amp^r gene and insertion of the kan^r gene (neomycin
phosphotransferase). The amp^r gene from the pUC backbone of V1J was removed by
digestion with SspI and Eam1105I restriction enzymes. The remaining plasmid was
purified by agarose gel electrophoresis, blunt-ended with T4 DNA polymerase, and
then treated with calf intestinal alkaline phosphatase. The commercially available
25 kan^r gene, derived from transposon 903 and contained within the pUC4K plasmid,
was excised using the PstI restriction enzyme, purified by agarose gel electrophoresis,
and blunt-ended with T4 DNA polymerase. This fragment was ligated with the V1J
backbone and plasmids with the kan^r gene in either orientation were derived which
were designated as V1Jneo #'s 1 and 3. Each of these plasmids was confirmed by
30 restriction enzyme digestion analysis, DNA sequencing of the junction regions, and
was shown to produce similar quantities of plasmid as V1J. Expression of
heterologous gene products was also comparable to V1J for these V1Jneo vectors.
V1Jneo#3, referred to as V1Jneo hereafter, was selected which contains the kan^r gene
in the same orientation as the amp^r gene in V1J as the expression construct and

provides resistance to neomycin, kanamycin and G418. The nucleotide sequence of V1Jneo is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGCT CTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
5 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CCCGACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCCGCTT ACATAACTTA CGGTAAATGG
10 CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
CATAGTAACG CCAATAGGGA CTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
15 CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
CGTCAATGGG AGTTTGGTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
CTCCGCCCA TTGACGCAA TGCGCGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
AGCTCGTTA GTGAACCGTC AGATGCCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGGGAT
20 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCA CCCCCCTTGGC
TTCTTATGCA TGCTATACTG TTTTGCGCTT GGGGTCTATA CACCCCGCT TCCTCATGTT
ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC
CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTGCC ACAACTCTCT
TTATTGGCTA TATGCCAATA CACTGTCCCT CAGAGACTGA CACGGACTCT GTATTTTAC
25 AGGATGGGTT CTCATTATT ATTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC
CCGCAGTTTT TATTAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTCCGG
ACATGGGCTC TTCTCGGTA CGGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC
CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACCA GTGGAGGCCA GACTTAGGCA
CAGCACGATG CCCACCAACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
30 TGAAAATGAG CTCGGGGAGC GGGCTTGCAC CGCTGACCGA TTTGGAAGAC TTAAGGCAGC
GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC
CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTTCCA TGGGTCTTTT
CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGGCC

CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCTT TCCTAATAAA
 ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATT C TATTCTGGGG GGTGGGGTGG
 GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
 GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTT CTCCTGGGCC AGAAAGAAC
 5 AGGCACATCC CCTTCTCTGT GACACACCCT GTCCACGCC C TGGTTCTTA GTTCCAGCCC
 CACTCATAGG ACACTCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC
 TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
 GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAAGAGGA GAGAAAATGC CTCCAACATG
 TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
 10 CGCTCGGTGCG TTCGGCTGCG GCGAGCGGT A TCAGCTCACT CAAAGGCGGT AATACGGTTA
 TCCACAGAAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCC GCAAAAGGCC
 AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTCCATA GGCTCCGCC CCCTGACGAG
 CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
 CAGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCCT GCCGCTTACC
 15 GGATACCTGT CCGCCTTCT CCCCTCGGG A AGCGTGGCGC TTTCTCAATG CTCACGCTGT
 AGGTATCTCA GTTCTGGTGA GGTCTGTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC
 GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA
 CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAAACAGGA TTAGCAGAGC GAGGTATGTA
 GGCAGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
 20 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTTTGA
 TCCGGAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG
 CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTT CTACGGGTC TGACGCTCAG
 TGGAACGAAA ACTCACGTTA AGGGATTTG GTCATGAGAT TATCAAAAG GATCTTCACC
 TAGATCCTTT TAAATTAAGA ATGAAGTTT AAATCAATCT AAAGTATATA TGAGTAAACT
 25 TGGTCTGACA GTTACCAATG CTTAACAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT
 CGTTCATCCA TAGTTGCCTG ACTCCGGGGG GGGGGGGCGC TGAGGTCTGC CTCGTGAAGA
 AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGGA
 GCCACGGTTG ATGAGAGCTT TGTGTAGGT GGACCAAGTT GTGATTTGA ACTTTGCTT
 TGCCACGGAA CGGTCTGCGT TGTCGGGAAG ATGGTGTGATC TGATCCTTC AACTCAGCAA
 30 AGTTCGATTT ATTCAACAAA GCGCCGTC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT
 TACAACCAAT TAACCAATTC TGATTAGAAA AACTCATCGA GCATCAAATG AACTGCAAT
 TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA GCGCTTCTG TAATGAAGGA
 GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGTC TGCGATTCCG
 ACTCGTCCAA CATCAATACA ACCTATTAAT TTCCCTCGT CAAAATAAG GTTATCAAGT

GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTCT
 TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC
 AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAA
 GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA
 5 ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTT CCCGGGGATC
 GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA
 GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
 CTACCTTCAG CATGTTTCAG AAACAACCTCT GGCGCATCGG GCTTCCCATA CAATCGATAG
 ATTGTCGCAC CTGATTGCC GACATTATCG CGAGCCCATT TATACCCATA TAAATCAGCA
 10 TCCATGTTGG AATTTAACG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA
 ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTCATGA TGATATATT
 TTATCTTGTG CAATGTAACA TCAGAGATT TGAGACACAA CGTGGCTTTC CCCCCCCCCC
 CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
 TAGAAAAATA AACAAATAGG GGTTCCGCC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
 15 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTT
 CGTC (SEQ ID NO:15).

V1Jns - The expression vector V1Jns was generated by adding an SfiI site to V1Jneo to facilitate integration studies. A commercially available 13 base pair SfiI linker (New England BioLabs) was added at the KpnI site within the BGH sequence 20 of the vector. V1Jneo was linearized with KpnI, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt SfiI linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated V1Jns. Expression of heterologous genes in V1Jns (with SfiI) was comparable to expression of the same genes in V1Jneo (with KpnI).

25 The nucleotide sequence of V1Jns is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
 CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 ACCATATGCG GTGTGAAATA CGGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
 30 CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTATAA TTGGCTCATG
 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
 GGGGTCACTTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
 CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC

TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCCGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
5 CGTCAATGGG AGTTTGTGTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
CTCCGCCCA TTGACGCAA TGCGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
AGCTCGTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCCGAT
TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGACTC TATAGGCACA CCCCTTGCG
10 TCTTATGCAT GCTATACTGT TTTTGGCTTG GGGCCTATAC ACCCCCCTT CCTTATGCTA
TAGGTGATGG TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
TATTGGTGCAC GATACTTCC ATTACTAATC CATAACATGG CTCTTGCCA CAACTATCTC
TATTGGCTAT ATGCCAATAC TCTGCTCTTC AGAGACTGAC ACGGACTCTG TATTTTTACA
GGATGGGGTC CCATTTATTA TTTACAAATT CACATATACA ACAACGCCGT CCCCCGTGCC
15 CGCAGTTTT ATTAAACATA GCGTGGGATC TCCACGCCAA TCTCGGGTAC GTGTTCCGGA
CATGGGCTCT TCTCCGGTAG CGGGGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCCTC
AGCGGCTCAT GGTGCGTCGG CAGCTCCTTG CTCTAACAG TGGAGGCCAG ACTTAGGCAC
AGCACAATGC CCACCACCCAC CAGTGTGCCG CACAAGGCCG TGGCGGTAGG GTATGTGTCT
GAAAATGAGC GTGGAGATTG GGCTCGCACG GCTGACGCCAG ATGGAAGACT TAAGGCAGCG
20 GCAGAAGAAG ATGCAGGCAG CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACCTCCC
GTTGCGGTGC TGTAAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGGC
CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT GGGCTTTTC
TGCAGTCACC GTCCTTAGAT CTGCTGTGCC TTCTAGTTGC CAGCCATCTG TTGTTTGCCC
CTCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC ACTGTCCCTT CCTAATAAAA
25 TGAGGAAATT GCATCGCATT GTCTGAGTAG GTGTCATTCT ATTCTGGGGG GTGGGGTGGG
GCAGGACAGC AAGGGGGAGG ATTGGGAAGA CAATAGCAGG CATGCTGGGG ATGCGGTGGG
CTCTATGGCC GCTGCGGCCA GGTGCTGAAG AATTGACCCG GTTCCCTCTG GGCCAGAAAG
AAGCAGGCAC ATCCCTCTCT CTGTGACACA CCCTGTCCAC GCCCCTGGTT CTTAGTTCCA
GCCCCACTCA TAGGACACTC ATAGCTCAGG AGGGCTCCG CTTCAATCCC ACCCGCTAAA
30 GTACTTGGAG CGGTCTCTCC CTCCCTCATC AGCCCACCAA ACCAAACCTA GCCTCCAAGA
GTGGGAAGAA ATTAAAGCAA GATAGGCTAT TAAGTGCAGA GGGAGAGAAA ATGCCTCCAA
CATGTGAGGA AGTAATGAGA GAAATCATAG AATTCTTCC GCTTCCCTCGC TCACTGACTC
GCTGCGCTCG GTCGTTCGGC TGCGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG
GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAA

GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCCTTTTC CATAGGCTCC GCCCCCCTGA
CGAGCATCAC AAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG
ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCCTCT CCTGTTCCGA CCCTGCCGCT
TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAAGCGTG GCGCTTCTC ATAGCTCACG
5 CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC
CCCCGTTCAAG CCCGACCGCT GCACCTTATC CGGTAACATAT CGTCTTGAGT CCAACCCGGT
AAGACACGAC TTATGCCAC TGCGAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA
TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC
AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC
10 TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTGTTTGCA ACCAGCAGAT
TACGCGCAGA AAAAAGGAT CTCAGAAGA TCCTTGATC TTTTCTACGG GGTCTGACGC
TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT
CACCTAGATC CTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA
AACTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT
15 ATTTCGTTCA TCCATAGTTG CCTGACTCGG GGGGGGGGG CGCTGAGGTC TGCGCTGTGA
AGAAGGTGTT GCTGACTCAT ACCAGGCCTG AATCGCCCCA TCATCCAGCC AGAAAGTGAG
GGAGCCACGG TTGATGAGAG CTTGTTGTA GGTGGACCAG TTGGTGTATT TGAACTTTG
CTTGCCACG GAACGGTCTG CGTTGTCGGG AAGATGCGTG ATCTGATCCT TCAACTCAGC
AAAAGTTCGA TTTATTCAAC AAAGCCGCCG TCCCGTCAAG TCAGCGTAAT GCTCTGCCAG
20 TGTTACAACC AATTAACCAA TTCTGATTAG AAAAACTCAT CGAGCATCAA ATGAAACTGC
AATTTATTCA TATCAGGATT ATCAATACCA TATTTTGAA AAAGCCGTTT CTGTAATGAA
GGAGAAAAT CACCGAGGCA GTTCCATAGG ATGGCAAGAT CCTGGTATCG GTCTGCGATT
CCGACTCGTC CAACATCAAT ACAACCTATT AATTTCCCCT CGTCAAAAT AAGGTTATCA
AGTGAGAAAT CACCATGAGT GACGACTGAA TCCGGTGAGA ATGGCAAAAG CTTATGCATT
25 TCTTCCAGA CTTGTTCAAC AGGCCAGCCA TTACGCTCGT CATCAAAATC ACTCGCATCA
ACCAAACCGT TATTCAATTG TGATTGCGCC TGAGCGAGAC GAAATACGCG ATCGCTGTAA
AAAGGACAAT TACAAACAGG AATCGAATGC AACCGGGCGA GGAACACTGC CAGCGCATCA
ACAATATTTT CACCTGAATC AGGATATTCT TCTAATACCT GGAATGCTGT TTTCCGGGG
ATCGCAGTGG TGAGTAACCA TGCAATCATCA GGAGTACGGA TAAAATGCTT GATGGTCGGA
30 AGAGGCATAA ATTCCGTCAG CCAGTTAGT CTGACCACATCT CATCTGTAAC ATCATTGGCA
ACGCTACCTT TGCCATGTTT CAGAAACAAAC TCTGGCGCAT CGGGCTTCCC ATACAATCGA
TAGATTGTCG CACCTGATTG CCCGACATTA TCGCGAGCCC ATTTATACCC ATATAAATCA
GCATCCCATGT TGGAATTAA TCGCGGCCCTC GAGCAAGACG TTTCCGTTG AATATGGCTC
ATAACACCCC TTGTATTACT GTTTATGTAA GCAGACAGTT TTATTGTTCA TGATGATATA

TTTTTATCTT GTGCAATGTA ACATCAGAGA TTTTGAGACA CAACGTGGCT TTCCCCCCCC
CCCCATTATT GAAGCATTAA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT
ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTC CCCGAAAAGT GCCACCTGAC
GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT CACGAGGCC
5 TTTCGTC (SEQ ID NO:16).

The underlined nucleotides of SEQ ID NO:16 represent the SfiI site introduced into the Kpn I site of V1Jneo.

V1Jns-tPA – The vaccine vector V1Jns-tPA was constructed in order to fuse an heterologous leader peptide sequence to the pol DNA constructs of the present invention. More specifically, the vaccine vector V1Jns was modified to include the human tissue-specific plasminogen activator (tPA) leader. As an exemplification, but by no means a limitation of generating a pol DNA construct comprising an amino-terminal leader sequence, plasmid V1Jneo was modified to include the human tissue-specific plasminogen activator (tPA) leader. Two synthetic complementary oligomers 10 were annealed and then ligated into V1Jneo which had been BglII digested. The sense and antisense oligomers were 5'-GATCACCATGGATGCAATGAAGAG
AGGGCTCTGCTGTGCTGCTGCTGTGGAGCAGTCTCGTTCGCCAG
CGA-3' (SEQ ID NO:17); and, 5'-GATCTCGCTGGCGAACGAAAGACTGCTCC
ACACAGCAGCAGCACACAGCAGAGCCCTCTTCATTGCATCCATGGT-3'
15 (SEQ ID NO:18). The Kozak sequence is underlined in the sense oligomer. These oligomers have overhanging bases compatible for ligation to BglII-cleaved sequences. After ligation the upstream BglII site is destroyed while the downstream BglII is retained for subsequent ligations. Both the junction sites as well as the entire tPA leader sequence were verified by DNA sequencing. Additionally, in order to conform 20 with V1Jns (=V1Jneo with an SfiI site), an SfiI restriction site was placed at the KpnI site within the BGH terminator region of V1Jneo-tPA by blunting the KpnI site with T4 DNA polymerase followed by ligation with an SfiI linker (catalogue #1138, New England Biolabs), resulting in V1Jns-tPA. This modification was verified by restriction digestion and agarose gel electrophoresis.
25 The V1Jns-tpa vector nucleotide sequence is as follows:
TCGC CGCTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CGGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
30

CTATTGGCCA TTGCATAACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
 GGGGTCACTTA GTTCATAGCC CATATATGGA GTTCCCGCGTT ACATAACTTA CGGTAAATGG
 CCCCGCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
 5 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
 TGCCCACCTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
 TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTCCTAC
 TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
 CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTTC CCAAGTCTCC ACCCCATTGA
 10 CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
 CTCCGCCCA TTGACGCCAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
 AGCTCGTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACCGCTGTT TTGACCTCCA
 TAGAACACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
 TCCCCGTGCC AAGACTGACG TAAGTACCGC CTATAGACTC TATAGGCACA CCCCTTTGGC
 15 TCTTATGCAT GCTATACTGT TTTTGGCTTG GGGCCTATAC ACCCCCCTT CCTTATGCTA
 TAGGTGATGG TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
 TATTGGTGAC GATACTTCC ATTACTAAC CATAACATGG CTCTTGCCA CAACTATCTC
 TATTGGCTAT ATGCCAATAC TCTGTCCTTC AGAGACTGAC ACGGACTCTG TATTTTTACA
 GGATGGGGTC CCATTTATTA TTTACAAATT CACATATACA ACAACGCCGT CCCCCGTGCC
 20 CGCAGTTTT ATTAAACATA GCGTGGGATC TCCACGCCAA TCTCGGGTAC GTGTTCCGGA
 CATGGGCTCT TCTCCGGTAG CGGGGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCCTCC
 AGCGGCTCAT GGTGGCTCGG CAGCTCCTTG CTCCCTAACAG TGGAGGCCAG ACTTAGGCAC
 AGCACAAATGC CCACCACCCAC CAGTGTGCCG CACAAGGCCG TGGCGGTAGG GTATGTGTCT
 GAAAATGAGC GTGGAGATTG GGCTCGCAG GCTGACGCCAG ATGGAAGACT TAAGGCAGCG
 25 GCAGAAGAAG ATGCAGGCAG CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC
 GTGCGGTGC TGTAAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGG
 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTCCAT GGGTCTTTTC
 TGCAGTCACC GTCCCTAGAT CACCATGGAT GCAATGAAGA GAGGGCTCTG CTGTGTGCTG
CTGCTGTGTG GAGCAGTCTT CGTTCGCCC AGCGAGATCT GCTGTGCCCTT CTAGTTGCCA
 30 GCCATCTGTT GTTTGCCCTT CCCCCGTGCC TTCCCTTGACC CTGGAAAGGTG CCACCTCCCAC
 TGTCCTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT CTGAGTAGGT GTCATTCTAT
 TCTGGGGGGT GGGGTGGGAGC AGGACAGCAA GGGGGAGGAT TGGGAAGACA ATAGCAGGCA
 TGCTGGGGAT GCGGTGGGCT CTATGGCCGC TGCGGCCAGG TGCTGAAGAA TTGACCCGGT
 TCCTCCTGGG CCAGAAAGAA GCAGGCACAT CCCCTCTCT GTGACACACC CTGTCCACCGC

CCCTGGTTCT TAGTTCCAGC CCCACTCATA GGACACTCAT AGCTCAGGAG GGCTCCGCCT
TCAATCCCAC CCGCTAAAGT ACTTGGAGCG GTCTCTCCCT CCCTCATCAG CCCACCAAAC
CAAACCTAGC CTCCAAGAGT GGGAAAGAAAT TAAAGCAAGA TAGGCTATTAGTGCAGAGG
GAGAGAAAAT GCCTCCAACA TGTGAGGAAG TAATGAGAGA AATCATAGAA TTTCTTCCGC
5 TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTCGGCTG CGGGCAGCGG TATCAGCTCA
CTCAAAGGCC GTAATACGGT TATCCACAGA ATCAGGGAT AACGCAGGAA AGAACATGTG
AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTCCA
TAGGCTCCGC CCCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA
CCCGACAGGA CTATAAAAGAT ACCAGGCCTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC
10 TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTCGG GAAGCGTGGC
GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTGGTG TAGTCGTTG GCTCCAAGCT
GGGCTGTGTG CACGAACCCC CGGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG
TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG
GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA
15 CGGCTACACT AGAAGAACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG
AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAAACCACC GCTGGTAGCG GTGGTTTTTT
TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT
TCTTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT
20 CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC
TATCTCAGCG ATCTGCTAT TTCGTTCATC CATAGTTGCC TGACTCGGGG GGGGGGGGGCG
CTGAGGTCTG CCTCGTGAAG AAGGTGTTGC TGACTCATAC CAGGCCTGAA TCGCCCCATC
ATCCAGCCAG AAAGTGAGGG AGCCACGGTT GATGAGAGCT TTGTTGTAGG TGGACCAGTT
GGTGTATTTG AACTTTGCT TTGCCACGGA ACGGTCTGCG TTGTCGGGAA GATGCGTGT
25 CTGATCCTTC AACTCAGCAA AAGTTCGATT TATTCAACAA AGCCGCCGTC CCGTCAAGTC
AGCGTAATGC TCTGCCAGTG TTACAACCAA TTAACCAATT CTGATTAGAA AAACTCATCG
AGCATCAAAT GAAACTGCAA TTTATTCTA TCAGGATTAT CAATACCATA TTTTGAAAAA
AGCCGTTCT GTAATGAAGG AGAAAAACTCA CCGAGGCAGT TCCATAGGAT GGCAAGATCC
TGGTATCGGT CTGCGATTC GACTCGTCCA ACATCAATAC AACCTATTAA TTTCCCTCG
30 TCAAAAATAA GGTTATCAAG TGAGAAATCA CCATGAGTGA CGACTGAATC CGGTGAGAAT
GGCAAAAGCT TATGCATTTT CTTCCAGACT TGTTCAACAG GCCAGCCATT ACGCTCGTCA
TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTGCG ATTGCGCCTG AGCGAGACGA
AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA TCGAATGCAA CCGGCGCAGG
AACACTGCCA CGCGCATCAAC AATATTTCAT CCTGAATCAG GATATTCTTC TAATACCTGG

AATGCTGTTT TCCCAGGGAT CGCAGTGGTG AGTAACCAG CATCATCAGG AGTACGGATA
AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTAGTCT GACCATCTCA
TCTGTAACAT CATTGGCAAC GCTACCTTG CCATGTTCA GAAACAACTC TGGCGCATCG
GGCTTCCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC GCGAGCCCAT
5 TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAAC GCGGGCTCGA GCAAGACGTT
TCCCCTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT TTATGTAAGC AGACAGTTTT
ATTGTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT TTGAGACACA
ACGTGGCTTT CCCCCCCCCC CCATTATTGA AGCATTATC AGGGTTATTG TCTCATGAGC
GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC
10 CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT
AGGCGTATCA CGAGGCCCTT TCGTC (SEQ ID NO:9).

V1R – Vaccine vector V1R was constructed to obtain a minimum-sized vaccine vector without unneeded DNA sequences, which still retained the overall optimized heterologous gene expression characteristics and high plasmid yields that 15 V1J and V1Jns afford. It was determined that (1) regions within the pUC backbone comprising the *E. coli* origin of replication could be removed without affecting plasmid yield from bacteria; (2) the 3'-region of the *kan^r* gene following the kanamycin open reading frame could be removed if a bacterial terminator was inserted in its place; and, (3) ~300 bp from the 3'- half of the BGH terminator could 20 be removed without affecting its regulatory function (following the original KpnI restriction enzyme site within the BGH element). V1R was constructed by using PCR to synthesize three segments of DNA from V1Jns representing the CMVintA promoter/BGH terminator, origin of replication, and kanamycin resistance elements, respectively. Restriction enzymes unique for each segment were added to each 25 segment end using the PCR oligomers: SspI and XhoI for CMVintA/BGH; EcoRV and BamHI for the *kan^r* gene; and, BclI and SalI for the *ori^r*. These enzyme sites were chosen because they allow directional ligation of each of the PCR-derived DNA segments with subsequent loss of each site: EcoRV and SspI leave blunt-ended DNAs which are compatible for ligation while BamHI and BclI leave complementary 30 overhangs as do SalI and XhoI. After obtaining these segments by PCR each segment was digested with the appropriate restriction enzymes indicated above and then ligated together in a single reaction mixture containing all three DNA segments. The 5'-end of the *ori^r* was designed to include the T2 rho independent terminator sequence that is normally found in this region so that it could provide termination

information for the kanamycin resistance gene. The ligated product was confirmed by restriction enzyme digestion (>8 enzymes) as well as by DNA sequencing of the ligation junctions. DNA plasmid yields and heterologous expression using viral genes within V1R appear similar to V1Jns. The net reduction in vector size achieved was
5 1346 bp (V1Jns = 4.86 kb; V1R = 3.52 kb). PCR oligomer sequences used to synthesize V1R (restriction enzyme sites are underlined and identified in brackets following sequence) are as follows: (1) 5'-GGTACAAAATATTGGCTATTGG
CCATTGCATACG-3' (SEQ ID NO:19) [SspI]; (2) 5'-CCACATCTCGAGGAAC
CGGGTCAATTCTTCAGCACC-3' (SEQ ID NO:20) [XhoI] (for CMVintA/BGH
10 segment); (3) 5'-GGTACAGAATATCGGAAGCCACGTGTG TCTAAAATC-3'
(SEQ ID NO:21) [EcoRV]; (4) 5'-CACATGGATCCGTAAT GCTCTGCCAGTGTT
ACAACC-3' (SEQ ID NO:2) [BamHI], (for kanamycin resistance gene segment) (5)
5'-GGTACATG ATCACGTAGAAAAGATCA AAGGATCTTCTTG-3' (SEQ ID
NO:23) [BclII]; (6) 5'-CCACATGTCGACCCGTAAA AAGGCCGCGTTGCTGG-3'
15 (SEQ ID NO:24); [SalI], (for *E. coli* origin of replication).

The nucleotide sequence of vector V1R is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
20 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
25 CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC
TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCCTAC
TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA
CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
30 CGTCAATGGG AGTTTGT~~TTT~~ GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA
CTCCGCCCCA TTGACGCAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
AGCTCGTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGAA CGGTGCATTG GAACGCGGAT
TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAAGGCCA CCCCCCTTGGC

TTCTTATGCA TGCTATACTG TTTTGCGCTT GGGGTCTATA CACCCCCGCT TCCTCATGTT
ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC
CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTGCC ACAACTCTCT
TTATTGGCTA TATGCCAATA CACTGTCCCT CAGAGACTGA CACGGACTCT GTATTTTAC
5 AGGATGGGGT CTCATTTATT ATTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC
CCGCAGTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTCCGG
ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC
CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACAA GTGGAGGCCA GACTTAGGCA
CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC
10 TGAAAATGAG CTCGGGAGC GGGCTTGCAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC
GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC
CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC
GCCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCCTTCCA TGGGTCTTTT
CTGCAGTCAC CGTCCTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGTTGCC
15 CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCTT TCCTAATAAA
ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTC TATTCTGGGG GGTGGGGTGG
GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG
GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAGAAC
AGGCACATCC CCTTCTCTGT GACACACCCCT GTCCACGCC CTCGGTTCTTA GTTCCAGCCC
20 CACTCATAGG ACACATCAG CTCAGGAGGG CTCCGCCCTTC AATCCCACCC GCTAAAGTAC
TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG
GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAAGAGGA GAGAAAATGC CTCCAACATG
TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
CGCTCGGTGCG TTGCGCTGCG GCGAGCGGT AACTACGGTTA
25 TCCACAGAAT CAGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCC GAAAAGGCC
AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCC CCCTGACGAG
CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC
CAGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCCT GCCGTTTAC
GGATACCTGT CCGCCTTCT CCCTCGGGA AGCGTGGCGC TTTCTCAATG CTCACGCTGT
30 AGGTATCTCA GTTCGGTGTAA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCC
GTTCAAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA
CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA
GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTTTGA

TCCGGCAAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG
CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTT CTACGGGGTC TGACGCTAG
TGGAACGAAA ACTCACCGTTA AGGGATTTG GTCATGAGAT TATCAAAAG GATCTTCACC
TAGATCCTTT TAAATTAAGA ATGAAGTTT AAATCAATCT AAAGTATATA TGAGTAAACT
5 TGGTCTGACA GTTACCAATG CTTAACAGT GAGGCACCTA TCTCAGCGAT CTGTCTATT
CGTTCATCCA TAGTTGCCCTG ACTCCGGGGG GGGGGGGCGC TGAGGTCTGC CTCGTGAAGA
AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGGA
GCCACGGTTG ATGAGAGCTT TGTGTTAGGT GGACCAAGTTG GTGATTTGA ACTTTGCTT
TGCCACGGAA CGGTCTGCGT TGTCGGAAG ATGCGTGATC TGATCCTCA ACTCAGCAA
10 AGTCGATTT ATTCAACAAA GCCGCCGTCC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT
TACAACCAAT TAACCAATT TGATTAGAAA AACTCATCGA GCATCAAATG AAACTGCAAAT
TTATTCAAT CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTCTG TAATGAAGGA
GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG
ACTCGTCAA CATCAATACA ACCTATTAAT TTCCCCTCGT CAAAAATAAG GTTATCAAGT
15 GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTTCT
TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC
AAACCGTTAT TCATTCTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAA
GGACAATTAC AACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA
ATATTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTT CCCGGGGATC
20 GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA
GGCATAAAATT CCGTCAGCCA GTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
CTACCTTGC CATGTTTCAG AAACAACCTCT GGCGCATCGG GCTTCCCATA CAATCGATAG
ATTGTCGCAC CTGATTGCC GACATTATCG CGAGCCATT TATACCCATA TAAATCAGCA
TCCATGTTGG AATTAAATCG CGGCCCTCGAG CAAGACGTT CCCGTTGAAT ATGGCTCATA
25 ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTA TTGTTCATGA TGATATATT
TTATCTTGTG CAATGTAACA TCAGAGATT TGAGACACAA CGTGGCTTTC CCCCCCCCCC
CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
TAGAAAAATA AACAAATAGG GGTCGGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTT
30 CGTC (SEQ ID NO:25).

EXAMPLE 2

Codon Optimized HIV-1 Pol and HIV-1 IA Pol Derivatives as DNA Vector Vaccines

Synthesis of WT-optpol and IA-opt-pol Gene - Construction of both genes were conducted by Midland Certified Reagent Company (Midland, TX) following

5 established strategies. Ten double stranded oligonucleotides, ranging from 159 to 340 bases long and encompassing the entire pol gene, were synthesized by solid state methods and cloned separately into pUC18. For the wt-pol gene, the fragments are as follows:

	<i>Bgl</i> II#1- <i>Ecl</i> 136II half site at 282	= pJS6A1-7
10	<i>Pml</i> II half site at #285 - <i>Ecl</i> 136II half site at #597	= pJS6B2-5
	<i>Ssp</i> I half site at #600 - <i>Ecl</i> 136II half site at #866	= pJS6C1-4
	<i>Sma</i> I half site at #869 - <i>Apa</i> I #1095	= pJS6D1-4
	<i>Apa</i> I #1095 - <i>Kpn</i> I #1296	= pJS6E1-4
	<i>Kpn</i> I #1296 - <i>Xcm</i> I #1636	= pJS6F1-5
15	<i>Xcm</i> I #1636 - <i>Nsi</i> I #1847	= pJS6G1-2
	<i>Nsi</i> I #1847 - <i>Bcl</i> II half site at #2174	= pJS6H1-14
	<i>Bcl</i> II half site at #2174 - <i>Sac</i> I #2333	= pJS6I1-2
	<i>Sac</i> I #2333 - <i>Bgl</i> III #2577	= pJS6J1-1

20 *Eco*RI and *Hind*III sequences were added upstream of each 5' end and downstream of each 3' end, respectively, to allow cloning into the *Eco*RI-*Hind*III sites of pUC18.

The next stage of the synthesis was to consolidate these cassettes into three roughly equal fragments (alpha, beta, gamma) and was performed as follows:

25 Alpha: The *Ssp*I-*Hind*III small fragment of pJS6C1-4 was transferred into the *Ecl*136II-*Hind*III sites of pJS6B2-5 to give pJS6BC1-1. Into the *Eco*RI-*Pml*II sites of this plasmid was inserted the *Eco*RI-*Ecl*136II small fragment of pJS6A1-7 to give pJS6α1-8.

30 Beta: The *Eco*RI-*Apa*I small fragment of pJS6D1-4 was inserted into the corresponding sites of pJS6E1-2 to give pJS6DE1-2. Also, the *Eco*RI-*Xcm*I small fragment of pJS6F1-5 was inserted into the corresponding sites of pJS6G1-2 to give pJS6FG1-1. Then the *Eco*RI-*Kpn*I small fragment of pJS6DE1-2 was inserted into the corresponding sites of pJS6FG1-1 to give pJS6β1-1.

Gamma: The *Sac*I-*Hind*III small fragment of pJS6J1-1 was inserted into the corresponding sites of pJS6I1-2 to give pJS6IJ1-1. This plasmid was propagated through *E. coli* SCS110 (*dam*-/*dcm*-) to permit subsequent cleavage at the *Bcl*II site.

The *BclI-HindIII* small fragment of the unmethylated pJS6IJ1-1 was inserted into the *BglII-HindIII* sites of pJS6H1-14 to give pJS6 χ 1-1.

The wt-pol alpha, beta, gamma were ligated into the entire sequence as follows:

- 5 The *EcoRI-Ecl136II* small fragment of pJS6 α 1-8 was inserted into the *EcoRI-SmaI* sites of pJS6 β 1-1 to give pJS6 $\alpha\beta$ 2-1.

Into the *NsiI-HindIII* sites of this plasmid was inserted the *NsiI-HindIII* small fragment of pJS6 χ 1-1 to give pUC18-wt-pol. This final plasmid was completely resequenced in both strands.

- 10 To construct the entire IA-pol gene, only 3 new small fragments were synthesized:

<i>PmlI</i> half site at #285 – <i>Ecl136II</i> half site at #597	= pJS7B1-1
<i>KpnI</i> #1296 – <i>XcmI</i> #1636	= pJS7F1-2
<i>NsiI</i> #1847 – <i>BglII</i> half site at #2174	= pJS7H1-5

- 15 These were then used in the same reconstruction strategy as described above to give pUC18-IA-pol.

Expression Vector Construction - pUC18-wt-pol and pUC18-IA-pol were digested with *BglII* in order to isolate fragments containing the entire pol genes. V1R, V1Jns, V1Jns-tpa (Shiver, et al., 1995, Immune responses to HIV gp120 elicited by

- 20 DNA vaccination. In *Vaccines 95* (eds. Chanock, R. M., Brown, F., Ginsberg, H.S., & Norrby, E.) @ pp. 95-98; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; see also Example Section 1) were digested with *BglII*. The cut vectors were then treated with calf intestinal alkaline phosphatase. Both wt-pol and IA-pol genes were ligated into cut V1R using T4 DNA ligase (16 °C, overnight).

- 25 Competent DH5 α cells were transformed with aliquots of the ligation mixtures.

Colonies were screened by restriction digestion of amplified plasmid isolates.

Following a similar strategy, the *BglII* fragment containing the IA-pol was subcloned into the *BglII* site of V1Jns. To ligate the IA-pol gene into V1Jns-tpa, the IA-pol gene was PCR-amplified from V1R-IA-pol using pfu polymerase and the following

- 30 pair of primers: 5'-GGTACAAGATCTCCGCCCCATCTCCCCATTGAGA-3' (SEQ ID NO:26), and 5'-CCACATAGATCTGCCGGGCTTAGTCCTCATC-3' (SEQ ID NO:27). The upstream primer was designed to remove the initiation met codon and place the pol gene in frame with the tpa leader coding sequence from V1Jns-tpa. The PCR product was purified from the agarose gel slab using Sigma

DNA Purification spin columns. The purified products were digested with *Bgl*II and subcloned into the *Bgl*II site of V1Jns-tpa.

- Results* - The codon humanized wt- and IA-pol genes were constructed via stepwise ligation of 10 synthetic dsDNA fragments (Ferretti, et al., 1986, *Proc. Natl. Acad. Sci. USA* 83: 599-603). For expression in mammalian systems, the IA-pol gene was subcloned into V1R, V1Jns, and V1Jns-tpa. All these vectors place the gene under the control of the human cytomegalovirus/intron A hybrid promoter (hCMVIA). The DNA sequence of the IA-pol gene and the expressed protein product are shown in Figure 2A-B. Subcloning into V1Jns-tpa attaches the leader sequence from human tissue-specific plasminogen activator (tpa) to the N-terminus of the IA-pol (Pennica, et al., 1983, *Nature* 301: 214-221) to allow secretion of the protein. The sequences of the tpa leader and the fusion junction are shown in Figure 3.

EXAMPLE 3

15 HIV-1 POL Vaccine - Rodent Studies

Materials - *E. coli* DH5 α strain, penicillin, streptomycin, ACK lysis buffer, hepes, L-glutamine, RPMI1640, and ultrapure CsCl were obtained from Gibco/BRL (Grand Island, NY). Fetal bovine serum (FBS) was purchased from Hyclone. Kanamycin, Tween 20, bovine serum albumin, hydrogen peroxide (30%), concentrated sulfuric acid, β -mercaptoethanol (β -ME), and concanavalin A were obtained from Sigma (St. Louis, MO). Female balb/c mice at 4-6 wks of age were obtained from Taconic Farms (Germantown, NY). 0.3-mL insulin syringes were purchased from Myoderm. 96-well flat bottomed Maxisorp plates were obtained form NUNC (Rochester, NY). HIV-1_{III}B RT p66 recombinant protein was obtained from Advanced Biotechnologies, Inc. (Columbia, MD). 20-mer peptides were synthesized by Research Genetics (Huntsville, AL). Horseradish peroxidase (HRP)-conjugated rabbit anti-mouse IgG1 was obtained from ZYMED (San Francisco, CA). 1,2-phenylenediamine dihydrochloride (OPD) tablets was obtained from DAKO (Norway). Purified rat anti-mouse IFN-gamma (IgG1, clone R4-6A2), biotin-conjugated rat anti-mouse IFN-gamma (IgG1, clone XMG 1.2), and strepavidin-alkaline phosphatase conjugate were purchased from PharMingen (San Diego, CA). 1-STEP NBT/BCIP dye was obtained from Pierce Chemicals (Rockford, IL). 96-well Multiscreen membrane plate was purchased from Millipore (France). Cell strainer was obtained from Becton-Dickinson (Franklin Lakes, NJ).

Plasmid Preparation - E. coli DH5 α cells expressing the pol plasmids were grown to saturation in LB broth supplemented with 100 ug/mL kanamycin. Plasmid were purified by standard CsCl method and solubilized in saline at concentrations greater than 5 mg/mL until further use.

- 5 *Vaccination* - The plasmids were prepared in phosphate-buffered saline and administered into balb/c by needle injection (28-1/2G insulin syringe) of 50 uL aliquot into each quad muscle. V1Jns-IApol was administered at 0.3, 3, 30 ug dose and for comparison, V1Jns-tpa-IApol was given at 30 ug dose. Immunizations were conducted at T=0 and T=8 wks (for select animals from the 30-ug dose cohorts).
- 10 *ELISA Assay* - At T=12 wks, blood samples were collected by making an incision of a tail vein and the serum separated. Anti-RT titers were obtained following standard secondary antibody-based ELISA. Briefly, Maxisorp plates were coated by overnight incubation with 100 uL of 1 ug/mL HIV-1 RT protein (in PBS). The plates were washed with PBS/0.05% Tween 20 and incubated for approx. 2h with
- 15 200 uL/well of blocking solution (PBS/0.05% tween/1% BSA). The blocking solution was decanted; 100 uL aliquot of serially diluted serum samples were added per well and incubated for 2 h at room temperature. The plates were washed and 100 uL of 1/1000-diluted HRP-rabbit anti-mouse IgG were added with 1 h incubation. The plates were washed thoroughly and soaked with 100 uL OPD/H₂O₂ solution for
- 20 15 min. The reaction was quenched by adding 100 uL of 0.5M H₂SO₄ per well. OD₄₉₂ readings were recorded.

- ELIsop - Spleens* were collected from 5 mice/cohort at T=13-14 wks and pooled into a tube of 8-mL R10 medium (RPMI1640, 10% FBS, 2mM L-glutamine, 100U/mL Penicillin, 100 u/mL streptomycin, 10 mM Hepes, 50 uM β -ME).
- 25 Multiscreen opaque plates were coated with 100 μ l/well of capture mAb (purified R4-6A2 diluted in PBS to 5 μ g/ml) at 4°C overnight. The plates were washed with PBS/Pen/Strep in hood and blocked with 200 μ l/well of complete R10 medium for 37°C for at least 2 hrs. The mouse spleens were ground on steel mesh, collected into 15ml tubes and centrifuged at 1200rpm for 10min. The pellet was treated in ACK buffer (4ml of lysis buffer per spleen) for 5min at room temperature to lyse red blood cells. The cell pellet was centrifuged as before, resuspended in K-medium (5ml per mouse spleen), filtered through a cell strainer and counted using a hemacytometer. Block medium was decanted from the plates and 100 μ l/well of cell samples (5.0x10e5 cells per well) plus antigens were added. Pol-specific CD4 $^+$ cells were stimulated

- using a mixture of previously identified two epitope-containing peptides (aa641-660, aa731-750). Antigen-specific CD8+ cells were stimulated using a pool of four peptide epitope-containing peptides (aa201-220, aa311-330, aa571-590, aa781-800) or with individual peptides. A final concentration of 4 ug/mL per peptide was used.
- 5 Each splenocyte sample is tested for IFN-gamma secretion by adding the mitogen, concanavalin A. Plates were incubated at 37°C, 5% CO₂ for 20-24 h. The plates were washed with PBS/0.05% Tween 20 and soaked with 100 uL/well of 5 ug/mL biotin-conjugated rat anti-mouse IFN- mAb (clone XMG1.2) at 4°C overnight. The plates were washed and soaked with 100 uL/well 1/2500 dilution of strepavidin-AP
- 10 (in PBS/0.005% Tween/5%FCS) for 30 min at 37 °C. Following a wash, spots were developed by incubating with 100μl/well 1-step NBT/BCIP for 6-10 min. The plates were washed with water and allowed to air dry. The number of spots in each wells were determined using a dissecting microscope and normalized to 10e6 cells.

Results - Single vaccination of balb/c mice with V1Jns-IAPol is able to induce antigen-specific antibody (Figure 4) and T cell (Figure 5) responses in a dose response manner. IFN-gamma secretion from splenocytes can be detected from 3 and 30 ug cohort following stimulation with pools of peptides that contain CD4+ and CD8+ T cell epitopes. These epitopes were identified by (1) screening 20-mer peptides that encompass the entire pol sequence and overlap by 10 amino acid for ability to stimulate IFN-gamma secretion from vaccinee splenocytes, and (2) determining the T cell type (CD4+ or CD8+) by depleting either population in an Elispot-assay. Addition of tpa leader sequence to the pol gene is able to induce comparable, if not slightly higher, frequencies of pol-specific CD4+ and CD8+ cells. A second immunization with either V1Jns-IAPol and V1Jns-tpa-IAPol resulted in effective boosting of the immune responses.

EXAMPLE 4

HTV-1 Pol Vaccine - Non Human Primate Studies

Materials - *E. coli* DH5α strain, penicillin, streptomycin, and ultrapure CsCl were obtained from Gibco/BRL (Grand Island, NY). Kanamycin and phytohemagglutinin (PHA-M) were obtained from Sigma (St. Louis, MO). 20-mer peptides were synthesized by SynPep (Dublin, CA) and Research Genetics (Huntsville, AL). 96-well Multiscreen Immobilon-P membrane plates were obtained from Millipore (France). Strepavidin-alkaline phosphatase conjugate were purchased

- form Pharmingen (San Diego, CA). 1-Step NBT/BCIP dye was obtained form Pierce Chemicals (Rockford, IL). Rat anti-human IFN-gamma mAb and biotin-conjugated anti-human IFN-gamma reagent were obtained from R&D Systems (Minneapolis, MN). Dynabeads M-450 anti-human CD4 were obtained from Dynal (Norway).
- 5 HIVp24 antigen assay was purchased from Coulter Corporation (Miami, FL). HIV-1_{IIIB} RT p66 recombinant protein was obtained from Advanced Biotechnologies, Inc. (Columbia, MD). Plastic 8 well strips/plates, flat bottom, Maxisorp, are obtained from NUNC (Rochester, NY). HIV+ human serum 9711234 was obtained from Biological Specialty Corp.
- 10 *Plasmid Preparation* - *E. coli* DH5 α cells expressing the pol plasmids were grown to saturation in LB supplemented with 100 ug/mL kanamycin. Plasmid were purified by standard CsCl method and solubilized in saline at concentrations greater than 5 mg/mL until further use.
- 15 *Vaccination* - Cohorts of 3 rhesus macaques (approx. 5-10 kg) were vaccinated with 5 mg dose of either V1Jns-IAPol or V1Jns-tpa-IAPol. The vaccine was administered by needle injection of two 0.5 mL aliquots of 5 mg/mL plasmid solution (in phosphate-buffered saline, pH 7.2) into both deltoid muscles. Prior to vaccination, the monkeys were chemically restraint with i.m. injection of 10 mg/kg ketamine. The animals were immunized 3x at 4 week intervals (T=0, 4, 8.wks).
- 20 *Sample Collection* - Blood samples were collected at T = 0, 4, 8, 12, 16, 18 wks; sera and PBMCs were isolated using established protocols.
- 25 *ELIsop Assay* - Immobilon-IP plates were coated with 100 uL/well of rat anti-human IFN-gamma mAb at 15 ug/mL at 4 °C overnight. The plates are then washed with PBS and block by adding 200 uL/well of R10 medium. 4x10e5 peripheral blood cells were plated per well and to each well, either media or one of the pol peptide pools (final concentration of 4 ug/mL per peptide) or PHA, a known mitogen, is added to a final volume of 100 uL. Duplicate wells were set up per sample per antigen and stimulation was performed for 20-24 h at 37 °C. The plates are then washed; biotinylated anti-human IFN-gamma reagent is added (0.1 ug/mL, 100 uL per well) and allowed to incubate for overnight at 4 °C. The plates are again washed and 100 uL of 1:2500 dilution of the strepavidin-alkaline phosphatase reagent (in PBS/0.005% Tween/5% FCS) is added and allowed to incubate for 2 h at ambient room temperature. After another wash, spots are developed by incubating with 100 uL/well of 1-step NBT/BCIP for 6-10 min. CD4- T cell depletion was performed by

adding 1 bead particle/10 cell of Dynabeads M450 anti-human CD4, prewashed with PBS, and incubating on the shaker at 4 °C for 30 min. The beads are fractionated magnetically and the unbound cells collected and quantified before plating onto the ELISpot assay plates (at 4x10e5 cells per well).

- 5 *CTL Assay* - Procedures for establishing bulk CTL culture with fresh or cryopreserved peripheral blood mononuclear cells (PBMC) are as follows. Twenty percent total PBMC were infected in 0.5 ml volume with recombinant vaccinia virus, Vac-tpaPol, respectively, at multiplicity of infection (moi) of 5 for 1 hr at 37°C, and then combined with the remaining PBMC sample. The cells were washed once in 10
- 10 ml R-10 medium, and plated in a 12 well plate at approximately 5 to 10 x 10⁶ cells/well in 4 ml R-10 medium. Recombinant human IL-7 was added to the culture at the concentration of 330 U/ml. Two or three days later, one milliliter of R-10 containing recombinant human IL-2 (100 U/ml) was added to each well. And twice weekly thereafter, two milliliters of cultured media were replaced with 2 ml fresh R-
- 15 10 medium with rhIL-2 (100 U/ml). The lymphocytes were cultured at 37°C in the presence of 5% CO₂ for approximately 2 weeks, and used in cytotoxicity assay as described below. The effector cells harvested from bulk CTL cultures were tested against autologous B lymphoid cell lines (BLCL) sensitized with peptide pools. To prepare for the peptide-sensitized targets, the BLCL cells were washed once with
- 20 R-10 medium, enumerated, and pulsed with peptide pool (about 4 to 8 µg/ml concentration for each individual peptide) in 1 ml volume overnight. A mock target was prepared by pulsing cells with peptide-free DMSO diluent to match the DMSO concentration in the peptide-pulsed targets. The cells were enumerated the next morning, and 1 x 10⁶ cells were resuspended in 0.5 ml R-10 medium. Five to ten
- 25 microliters of Na⁵¹CrO₄ were added to the tubes at the same time, and the cells were incubated for 1 to 2 hr 37°C. The cells were then washed 3 times and resuspended at 5x10⁴ cells/ml in R-10 medium to be used as target cells. The cultured lymphocytes were plated with target cells at designated effector to target (E:T) ratios in triplicates in 96-well plates, and incubated at 37°C for 4 hours in the presence of 5% CO₂. A
- 30 sample of 30 µl supernatant from each well of cell mixture was harvested onto a well of a Lumaplate-96 (Packard Instrument, Meriden, CT), and the plate was allowed to air dry overnight. The amount of ⁵¹Cr in the well was determined through beta-particle emission, using a plate counter from Packard Instrument. The percentage of specific lysis was calculated using the formula as: % specific lysis = (E-S) / (M-S).

The symbol E represents the average cpm released from target cells in the presence of effector cells, S is the spontaneous cpm released in the presence of medium only, and M is the maximum cpm released in the presence of 2% Triton X-100.

ELISA Assay - The pol-specific antibodies in the monkeys were measured in a competitive RT EIA assay, wherein sample activity is determined by the ability to block RT antigen from binding to coating antibody on the plate well. Briefly, Maxisorp plates were coated with saturating amounts of pol positive human serum (97111234). 250 uL of each sample is incubated with 15 uL of 266 ng/mL RT recombinant protein (in RCM 563, 1% BSA, 0.1% tween, 0.1% NaN₃) and 20 uL of lysis buffer (Coulter p24 antigen assay kit) for 15 min at room temperature. Similar mixtures are prepared using serially diluted samples of a standard and a negative control which defines maximum RT binding. 200 uL/well of each sample and standard were added to the washed plate and the plate incubated 16-24 h at room temperature. Bound RT is quantified following the procedures described in Coulter p24 assay kit and reported in milliMerck units per mL arbitrarily defined by the chosen standard.

Results - Repeated vaccinations with V1Jns-IApol induced in 1 of 3 monkeys (94R033) significant levels of antigen-specific T cell activation (Figure 6A-C and Table 2) and CTL killing of peptide-pulsed autologous cells (Figure 7A-B). A significant CD8+ component to the T cell responses in this animal was confirmed by peptide-stimulation of CD4-depleted PBMCs in an ELIspot assay (Table 2).

Immunization with V1Jns-tpa-IApol produced T cell responses from all 3 vaccinees (Figures 6A-C, Figure 7A-B; Table 2). Two (920078, 94R028) exhibited bulk CTL activity and detectable CD8+ components as measured by Elispot analyses of CD4-depleted PBMCs. For the third monkey (920073), the activated T cells were largely CD4+ (Table 2). Table 3 shows the time course data on the frequency of IFN-gamma secreting cells (SFC/million cells) upon antigen-specific stimulation for monkeys vaccinated 3x with either V1Jns-IApol or V1Jns-tpa-IApol (5 mg dose). At T=18 wks, CD4-cell depletion were performed; the reported values are the number of spots per million of fractionated cells and are not corrected for the resultant enrichment of CD8+ T cells. PBMCs were stimulated with peptide pools that represent either IA pol protein (mpol-1, mpol-2) or wt Pol (wtpol-1, wtpol-2).

TABLE 2

Vaccine	Animl No.	Antigen	T=0 Wk	T=4 Wk	T=8 Wk	T=12 Wk	T=18 Wk	
			Dose 1	Dose 2	Dose 3			
VIJns-1Apd 5 mgs	94R008	medium	1	15	6	11	11	11
		mpd-1	3	69	28	61	20	15
		mpd-2	0	25	21	19	28	16
		wmpd-1		49	20	53	18	
		wmpd-2		34	24	24	19	
	94R013	medium	0	14	6	9	18	11
		mpd-1	0	9	63	25	34	9
		mpd-2	1	15	24	36	24	15
		wmpd-1		9	50	33	18	
		wmpd-2		6	21	29	25	
	94R033	medium	4	15	11	14	13	8
		mpd-1	3	29	86	51	41	24
		mpd-2	0	24	25	43	59	64
		wmpd-1		30	38	60	53	
		wmpd-2		48	46	86	61	
VIJns-tpolApd 5 mgs	920078	medium	0	24	13	11	14	11
		mpd-1	3	110	120	119	155	11
		mpd-2	1	221	130	561	289	145
		wmpd-1		115	53	70	116	
		wmpd-2		218	204	490	194	
	920073	medium	0	13	3	15	15	6
		mpd-1	0	36	51	113	90	14
		mpd-2	0	29	16	83	115	34
		wmpd-1		20	35	100	74	
		wmpd-2		25	16	79	61	
	94R028	medium	0	18	11	18	19	9
		mpd-1	1	30	24	29	30	28
		mpd-2	1	24	23	66	59	95
		wmpd-1		23	25	34	29	
		wmpd-2		26	28	71	40	
	Näve	medium	1	19	3	38	9	4
		mpd-1	0	24	11	25	4	6
		mpd-2	1	24	5	28	6	5
		wmpd-1		18	13	20	6	
		wmpd-2		23	14	33	14	

For the Elispot assay, antigen specific stimulation were performed by using pools of 20-mer peptide pools based on the vaccine sequence. The vaccine pol sequence differs from the wild-type HIV-1 sequence by 9 point mutations, thereby affecting 16 of the 20-mer peptides in the pool. Comparable responses were observed
 5 in the vaccinees when these peptides are replaced with those using the wild-type sequences.

Four of the vaccinees gave anti-RT titers above background after 3 dosages of the plasmids (Table 2).

10

TABLE 3

Anti-RT levels in Rhesus Macaques Vaccinated 3x (4 week intervals) with 5 mgs of V1Jns-IApol or V1Jns-tpa-IApol expressed in mMU/mL.

Vaccine/Monkey	T=0 Wk	T=4	T=8	T=12	T=16
	DOSE 1	DOSE 2	DOSE 3		
V1Jns-IApol, 5 mg					
94R008	ND	<10	<10	15	14
94R013	ND	<10	<10	<10	<10
94R033	ND	<10	<10	25	19
V1Jns-tpa-IApol, 5 mg					
920078	ND	<10	<10	35	17
920073	ND	<10	<10	<10	<10
94R028	ND	<10	<10	20	63

15

EXAMPLE 5

Effect of Codon Optimization on In Vivo Expression and
 Cellular Immune Response of wt-pol

Materials and Methods - Extraction of virus-derived pol gene - The gene for RT-IN (wt-pol; a non-codon optimized wild type pol gene derived directly from the HIV IIIB genome) was extracted and amplified from the HIV IIIB genome using two primers, 5'-CAG GCG AGA TCT ACC ATG GCC CCC ATT AGC CCT ATT GAG ACT GTA-3' (SEQ ID NO:29) and 5'-CAG GCG AGA TCT GCC CGG GCT TTA ATC CTC ATC CTG TCT ACT TGC CAC-3' (SEQ ID NO:30), containing *Bgl*II sites.
 20 The reaction contained 200 nmol of each primer, 2.5 U of pfu Turbo DNA polymerase (Stratagene, La Jolla, CA), 0.2 mM of each dNTPs, and the template DNA in 10mM KCl, 10mM (NH₄)₂SO₄, 20mM Tris-HCl pH 8.75, 2mM MgSO₄, 0.1% TritonX-100, 0.1mg/ml bovine serum albumin (BSA). Thermocycling
 25

conditions were as follows: 20 cycles of 1 min at 95 °C, 1 min at 56 °C, and 4 mins at 72 °C with 15-min capping at 72 °C. The digested PCR fragment was subcloned into the *Bgl*II site of the expression plasmid V1Jns (Shiver, et al., 1995, Immune responses to HIV gp120 elicited by DNA vaccination. In Chanock, R. M., Brown, F., Ginsberg, 5 H. S., and Norrby, E. (Eds.) *Vaccines 95*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 95-98; see also Example section 1 herein) expression plasmid following similar procedures as described above. The ligation mixtures were then used to transform competent *E. coli* DH5 cells and screened by PCR amplification of individual colonies. Sequence of the entire gene insert was 10 confirmed. All plasmid constructs for animal immunization were purified by CsCl method (Sambrook, et al., 1989, Fritsch and Maniatis, T. (Eds) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor).

In vitro expression in mammalian cells - 1.5×10^6 293 cells were transfected with 1 or 10 µg of V1R-wt-pol (codon optimized) and V1Jns-wt-pol (virus derived) 15 using the Cell Pfect kit and incubated for 48 h at 37 °C, 5% CO₂, 90% humidity. Supernatants and cell lysates were prepared and assayed for protein content using Pierce Protein Assay reagent (Rockford, IL). Aliquots containing equal amounts of total protein were loaded unto 10-20% Tris glycine gel (Novex, San Diego, CA) along with the appropriate molecular weight markers. The pol product was detected using 20 anti-serum from a seropositive patient (Scripps Clinic, San Diego, CA) diluted 1:1000 and the bands developed using goat anti-human IgG-HRP (Bethyl, Montgomery, TX) at 1:2000 dilution and standard ECL reagent kit (Pharmacia LKB Biotechnology, Uppsala, Sweden).

Ultrasensitive RT activity assay of pol constructs - RT activities from codon 25 optimized wt-pol and IA pol plasmids were analyzed by the Product-Enhanced Reverse Transcriptase (PERT) assay using Perkin Elmer 7700, Taqman technology (Arnold, et al., 1999, One-step fluorescent probe product-enhanced reverse transcriptase assay. In McClelland, M., Pardee, A. (Eds.) *Expression genetics: accelerated and high-throughput methods*. Biotechniques Books, Natick, MA, pp. 30 201-210). Background levels for this assay were determined using 1:100,000 dilution of lysates from mock (chemical treatment only, no vector) transfected 293 cells. This background range is set as RT/reaction tube of 0.00 to 56.28 which is taken from the mean value of 13.80 +/- 3 standard deviations (sd=14.16). Any individual value >56.28 would be considered positive for PERT assay. Cells lysates were prepared

similarly for the following samples: mock transfection with empty V1Jns vector; no vector control; transfection with V1Jns-tpa-pol (codon optimized); and transfection with V1Jns-IApol (codon optimized). Samples were serially diluted to 1:100,000 in PERT buffer and 24 replicates for each sample at this dilution were assayed for RT activity.

- 5 *Rodent immunization with optimized and virus-derived pol plasmids* - To compare the immunogenic properties of wt-pol (codon optimized) and virus-derived pol gene, cohorts of BALB/c mice (N=10) were vaccinated with 1 µg, 10 µg, and 100 µg doses of V1R-wt-pol (codon optimized) and V1Jns-wt-pol plasmid (virus derived).
- 10 At 5 weeks post dose 1, 5 of 10 mice per cohort were boosted with the same dose of plasmid they initially received. In all cases, the vaccines were suspended or diluted in 6 mM sodium phosphate, 150 mM sodium chloride, pH 7.2, and the total dose was injected to both quadricep muscles in 50 µL aliquots using a 0.3-mL insulin syringe with 28-1/2G needles (Becton-Dickinson, Franklin Lakes, NJ).
- 15 *Anti-RT ELISA* - Anti-RT titers were obtained following standard secondary antibody-based ELISA. Maxisorp plates (NUNC, Rochester, NY) were coated by overnight incubation with 100 µL of 1 µg /mL HIV-1 RT protein (Advanced Biotechnologies, Columbia, MD) in PBS. The plates were washed with PBS/0.05% Tween 20 using Titertek MAP instrument (Hunstville, AL) and incubated for approximately 2h with 200 µL/well of blocking solution (PBS/0.05% tween/1% BSA). The blocking solution was decanted; 100 µL aliquot of serially diluted serum samples were added per well and incubated for 2 h at room temperature. An initial dilution of 100-fold is performed followed by 4-fold serial dilution. The plates were washed and 100 µL of 1/1000-diluted HRP-rabbit anti-mouse IgG (ZYMED, San Francisco, CA) were added with 1 h incubation. The plates were washed thoroughly and soaked with 100 µL 1,2-phenylenediamine dihydrochloride/hydrogen peroxide (DAKO, Norway) solution for 15 min. The reaction was quenched by adding 100 µL of 0.5M H₂SO₄ per well. OD₄₉₂ readings were recorded using Titertek Multiskan MCC/340 with S20 stacker. Endpoint titers were defined as the highest serum dilution that resulted in an absorbance value of greater than or equal to 0.1 OD₄₉₂ (2.5 times the background value):
- 20 25 30

ELIspot assay - Antigen-specific INF γ -secreting cells from mouse spleens were detected using the ELIsplot assay (Miyahira, et al., 1995, Quantification of antigen specific CD8 $^{+}$ T cells using an ELISPOT assay. *J. Immunol. Methods* 1995,

181, 45-54). Typically, spleens were collected from 3-5 mice/cohort and pooled into a tube of 8-mL complete RPMI media (RPMI1640, 10% FBS, 2mM L-glutamine, 100U/mL Penicillin, 100 u/mL streptomycin, 10 mM Hepes, 50 uM β -ME).

Multiscreen opaque plates (Millipore, France) were coated with 100 μ L/well of 5 μ g/mL purified rat anti-mouse IFN- γ IgG1, clone R4-6A2 (Pharmingen, San Diego, CA), in PBS at 4°C overnight. The plates were washed with PBS/penicillin/streptomycin in hood and blocked with 200 μ L/well of complete RPMI media for 37 °C for at least 2 h. The mouse spleens were ground on steel mesh, collected into 15ml tubes and centrifuged at 1200rpm for 10 min. The pellet was treated with 4 mL ACK buffer (Gibco/BRL) for 5 min at room temperature to lyse red blood cells. The cell pellet was centrifuged as before, resuspended in complete RPMI media (5 ml per mouse spleen), filtered through a cell strainer and counted using a hemacytometer. Block media was decanted from the plates and to each well, 100 μ L of cell samples (5×10^5 cells per well) and 100 μ L of the antigen solution were added. To the control well, 100 μ L of the media were added; for specific responses, peptide pools containing either CD4 $^+$ or CD8 $^+$ epitopes were added. In all cases, a final concentration of 4 μ g/mL per peptide was used. Each sample/antigen mixture were performed in triplicate wells. Plates were incubated at 37°C, 5% CO₂, 90% humidity for 20-24 h. The plates were washed with PBS/0.05% Tween 20 and incubated with 100 μ L/well of 1.25 μ g/mL biotin-conjugated rat anti-mouse IFN- γ mAb, clone XMG1.2 (Pharmingen) at 4°C overnight. The plates were washed and incubated with 100 μ L/well 1/2500 dilution of strepavidin-alkaline phosphatase conjugate (Pharmingen) in PBS/0.005% Tween/5% FBS for 30 min at 37 °C. Following a wash, spots were developed by incubating with 100 μ l/well 1-step NBT/BCIP (Pierce Chemicals) for 6-10 min. The plates were washed with water and allowed to air dry. The number of spots in each well was determined using a dissecting microscope and the data normalized to 10^6 cell input.

Results - In vitro expression of Pol in mammalian cells - Heterologous expression of the optimized wt or IA pol genes (V1R-wt-pol (codon optimized), V1Jns-IApol (codon optimized), V1Jns-tpa-IApol (codon optimized)) in 293 cells (Figure 8) yielded a single polypeptide of correct approximate molecular size (90-kDa) for the RT-IN fusion product. In contrast, no expression could be detected by transfecting cells with 1 and 10 μ g of the V1Jns-wt-pol, which bears the virus-derived pol.

Ultrasensitive RT assay of cells transfected with Pol constructs - Table 4
 summarizes the levels of polymerase activity from mock (vector only) control, IApol (codon optimized) and wt-pol plasmids (codon optimized). Results indicate that the wild-type POL transfected cells contained RT activity approximately 4-5 logs higher
 5 than the 293 cell only baseline values. Mock transfected cells contained activity no higher than baseline values. The RT activity from opt-IApol-transfected cells was also found to be no different than baseline values; no individual reaction tube resulted in RT activity higher than the established cut-off value of 56.

10

Table 4

Sample	Avg. RT/tube	Standard deviation	Minimum	Maximum
Vector only	16.25	18.52	0.0	42.99
IApol (codon optimized)	2.99	8.01	0.0	35.20
Wt-pol (codon optimized)	126147	21338	68973	152007

*Comparative immunogenicity of optimized and virus-derived pol plasmid - To compare the *in vivo* potencies of both constructs, BALB/c mice (N=10 per group)
 15 were vaccinated with escalating doses (1, 10, 100 µg) of either V1Jns-wt-pol (virus derived) or V1R-wt-pol (codon optimized). At 5 wks post dose 1, 5 of 10 animals were randomly boosted with the same vaccine and dose they received initially.
 Figure 9 shows the geometric mean titers of the BALB/c cohorts determined at 2 wks past boost. No significant anti-RT titers can be observed from animals immunized
 20 with one or two doses of the wt-pol plasmid (virus derived). In contrast, animals vaccinated with the humanized gene construct gave cohort anti-RT titers (>1000) significantly above background levels at doses above 10 ug. The responses seen at 10 and 100 ug dose of V1R-wt-pol (codon optimized) were boosted approximately 10-fold with a second immunization, reaching titers as high as 10⁶.
 25 Spleens from all mice in each of the cohorts were collected to be analyzed for IFN-γ secretion following stimulation with mixtures of either CD4+ peptide epitopes or CD8+ peptide epitopes. The results are shown in Figure 10. All wt-pol vaccinees did*

not show any significant cellular response above the background controls. In contrast, strong antigen-stimulated IFN- γ secretion were observed in a dose-responsive manner from animals vaccinated with one or two doses of 10 or more μ g of the wt-pol (codon optimized) construct.

- 5 The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

10

WHAT IS CLAIMED IS:

1. A pharmaceutically acceptable DNA vaccine composition, which comprises:
 - (a) a DNA expression vector; and,
 - 5 (b) a DNA molecule containing a codon optimized open reading frame encoding a Pol protein or inactivated Pol derivative thereof, wherein upon administration of the DNA vaccine to a host the Pol protein or inactivated Pol derivative is expressed and generates a cellular immune response against HIV-1 infection.
- 10 2. The DNA vaccine of claim 1 wherein the DNA molecule encodes wild type Pol.
- 15 3. The DNA vaccine of claim 2 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:1.
4. The DNA vaccine of claim 3 which is V1Jns-wt-pol.
5. The DNA vaccine of claim 1 wherein the DNA molecule encodes an inactivated Pol derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.
- 20 6. The DNA vaccine of claim 5 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:5
- 25 7. The DNA vaccine of claim 6 which is V1Jns-tPA-wt-pol.
8. The DNA vaccine of claim 1 wherein the inactivated Pol protein contains at least one amino acid modification within each region of the Pol protein responsible for reverse transcriptase activity, RNase H activity and integrase activity, such that the inactivated Pol protein shows no substantial reverse transcriptase activity, RNase H activity and integrase activity.

9. The DNA vaccine of claim 8 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:3
10. The DNA vaccine of claim 9 which is V1Jns-IAPol.
5
11. The DNA vaccine of claim 8 wherein the DNA molecule encodes an inactivated Pol derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.
- 10 12. The DNA vaccine of claim 11 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:7.
13. The DNA vaccine of claim 7 which is V1Jns-tPA-IAPol.
- 15 14. A method for inducing an immune response against infection or disease caused by virulent strains of HIV which comprises administering into the tissue of a mammalian host a pharmaceutically acceptable DNA vaccine composition which comprises a DNA expression vector and a DNA molecule containing a codon optimized open reading frame encoding a Pol protein or inactivated Pol derivative
20 thereof, wherein upon administration of the DNA vaccine to the vertebrate host the Pol protein or inactivated Pol derivative is expressed and generates the immune response.
15. The method of claim 16 wherein the mammalian host is a human.
25
16. The method of claim 17 wherein the DNA vaccine is selected from the group consisting of V1Jns-WTPol, V1Jns-tPA-WTPol, V1Jns-IAPol and V1Jns-tPA-IAPol.
- 30 17. A substantially purified protein which comprises an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8.

1/11

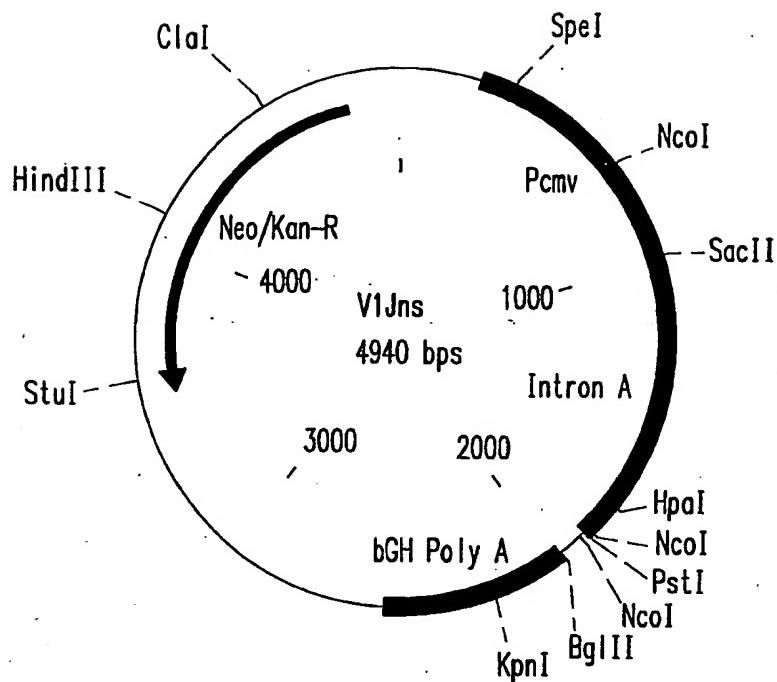


FIG. 1A

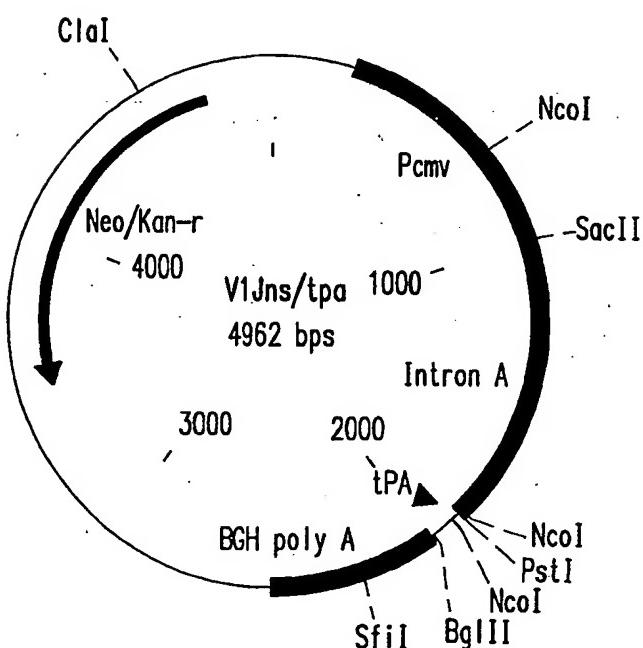


FIG. 1B

3/11

TGACTGAGGTGATCCCCCTGACTGAGGAGCCTGAGCTGGAGCTGGCTGACAACAGGGAGATCCTGAAGGAGCCTGTGCAT
 EuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis
 300 310

GGGGTGTACTATGACCCCTCAAGAACCTGAGACTGGCAAGTATGCCAGGATGAGGGGGCCCACACCAATGATGTCAACCACCTGA
 GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTy
 320 330 340

CCAGGAGCCCTTCAGAACCTGAAAGACTGGCAAGTATGCCAGGATGAGGGGGCCCACACCAATGATGTCAACCACCTGA
 rGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
 350 360 370

CTGAGGCTGTGCAGAACATCACCACTGAGTCCATTGTGATCTGGGCAAGACCCCCAAGTTCAAGCTGCCATCCAGAAG
 hrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLys
 380 390

GAGACCTGGGAGACCTGGTGGACTGAGTACTGGCAGGCCACCTGGATCCCTGAGTGGGAGTTTGTAACACCCCCCCCCCT
 GluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLe
 400 410 420

GGTGAAGCTGTGGTACCACTGGAGAAGGAGGCCATTGTGGGGCTGAGACCTCTATGTGGCTGGGCTGCCAACAGGG
 uValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAlaGlyAlaAlaAsnArgG
 430 440 450

AGACCAAGCTGGCAAGGCTGGCTATGTGACCAACAGGGCCAGGCAGAACGTTGGTACCCCTGACTGACACCACCAACCAG
 IuThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGin
 460 470

AAGACTGCCCTCCAGGCCATCTACCTGGCCCTCCAGGACTCTGGCCTGGAGGTGAAACATTGTGACTGCCCTCCAGTATGC
 LysThrAlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAlaSerGlnTyrAl
 480 490 500

CCTGGGCATCATCCAGGCCAGCCTGATCAGTCTGAGTCTGAGCTGGTCAACCAGATCATTGAGCAGCTGATCAAGAAGG
 oLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysG
 510 520 530

AGAAGGTGTACCTGGCTGGTGCCTGCCACAAGGGCATTGGGGCAATGAGCAGCTGACAAGCTGGTGTCTGCTGGC
 lLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGly
 540 550

ATCAGGAAGGTGCTGTTCTGGATGGCATTGACAAGGCCAGGATGAGCATGAGAACTACCAACTCCAAGCTGGAGGGCTAT
 IleArgLysValLeuPheLeuAspGlyIleAspLysAlaGlnAspGluHisGluLysTyrHisSerAsnTrpArgAlaMe
 560 570 580

FIG.2B

SUBSTITUTE SHEET (RULE 26)

4/11

GGCCTCTGACTTCAACCTGCCCTGTGGCTAAGGAGATTGTGCCCTCTGACAAGTCCAGCTGAACGGGAGG
 tAlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGlu
 590 600 610

CCATGCATGGGAGGTGGACTGCTCCCCTGCCATCTGGCAGCTGGCCTGCACCCACCTGGAGGGCAAGGTGATCCTGGTG
 lMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAlaCysThrHisLeuGluGlyLysValIleLeuVal
 620 630

GCTGTCATGTGCCCTCCGGCTACATTGAGGCTGAGGTATCCCTGCTGAGACAGGCCAGGAGACTGCCACTTCTGCT
 AlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLe
 640 650 660

GAAGCTGGCTGGCAGGTGGCTGTGAAGACCATCCACACTGCCAATGGCTCCAACTTCACTGGGCCACAGTGAGGGCTG
 uLysLeuAlaGlyArgTrpProValLysThrIleHisThrAlaAsnGlySerAsnPheThrGlyAlaThrValArgAlaA
 670 680 690

CCTGCTGGCTGGCATCAAGCAGGAGTTGGCATCCCTACAACCCCCAGTCCCAGGGGTGGTGGCCTCCATGAAC
 lCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGlnGlyValValAlaSerMetAsn
 700 710

AAGGAGCTGAAGAACATCATTGGCAGGTAGGGACCAGGCTGAGCACCTGAAGACAGCTGTGCAGATGGCTGTGTTCAT
 LysGluLeuLysLysIleIleGlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAlaValPhell
 720 730 740

CCACAACCAAGAGGAAGGGGGCATGGGGCTACTCCGCTGGGAGAGGATTGTCACATCATTGCCACAGACATCC
 eHisAsnPheLysArgLysGlyGlyIleGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIleG
 750 760 770

AGACCAAGGAGCTCCAGAACGATCACCAAGATCCAGAACTTCAGGGTGTACTACAGGACTCCAGGAACCCCCCTGTGG
 InThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArgAspSerArgAsnProLeuTrp
 780 790

AAGGGCCCTGCCAAGCTGCTGTGCAAGGGGAGGGGGCTGTGGTATCCAGGACAACCTGACATCAAGGTGGTGGCCAG
 LysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysValValProAr
 800 810 820

GAGGAAGGCCAAGATCATCAGGGACTATGGCAAGCAGATGGCTGGGATGACTGTGTGGCCTCCAGGCAGGATGAGGACT
 gArgLysAlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAspx
 830 840 850

AAAGCCCCGGCAGATCT (SEQ ID NO: 3)
 Xx *Bg*II

FIG.2C

5/11

GATCACCATTGCAATGAAGAGAGGCCCTGGCTGCTGCTGAGGAGTCCTCGTTCCG MetAspAlaMetLysArgIleLeuCysCysValLeuLeuCysGlyAlaValPheValSerP	-25	-10	
CCAGGGAGATCTCGCCCCATCTCCCCATCTGAGACTGTGAGCTGAACTGCTGGCATGGATGCC ProSerGluIleSerAlaProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly	-1	2	
			(within SEQ ID NO: 7) (within SEQ ID NO: 8)

FIG. 3

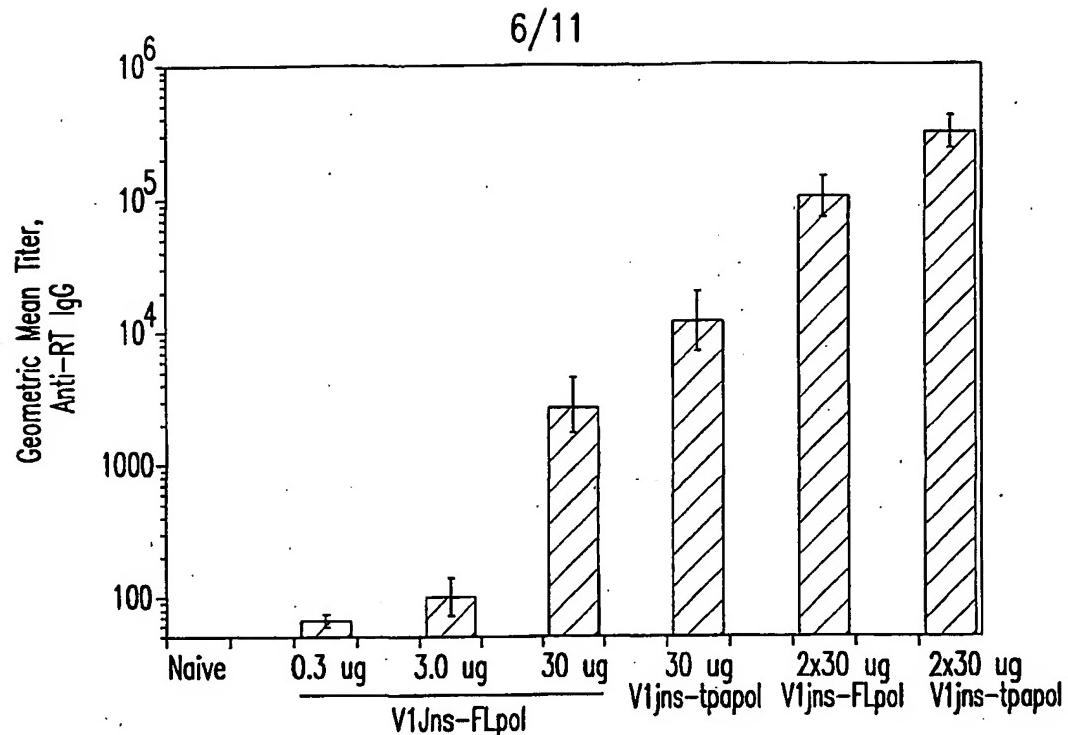


FIG.4

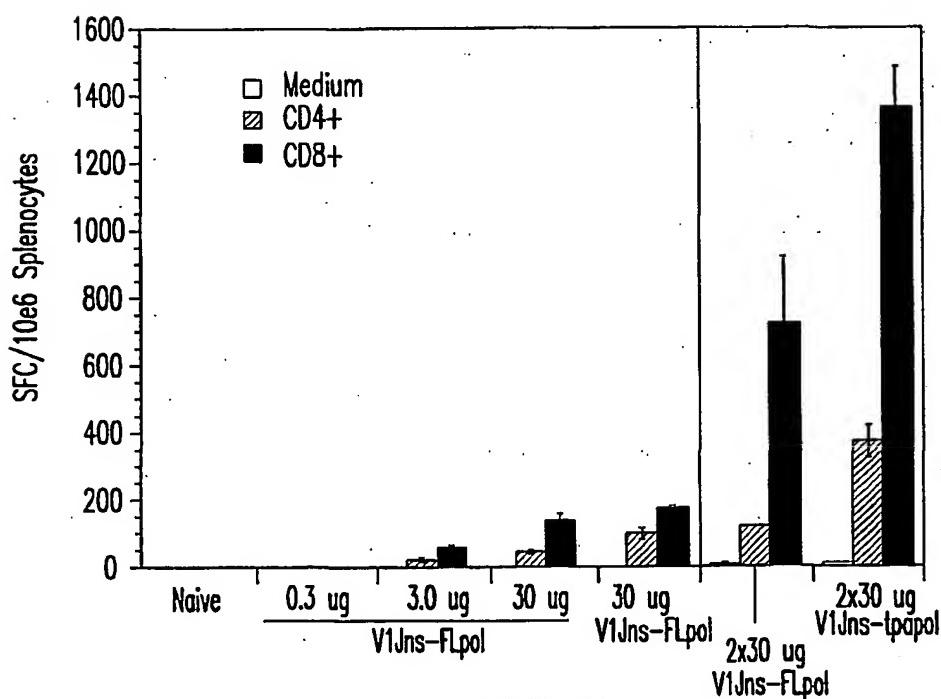


FIG.5

7/11

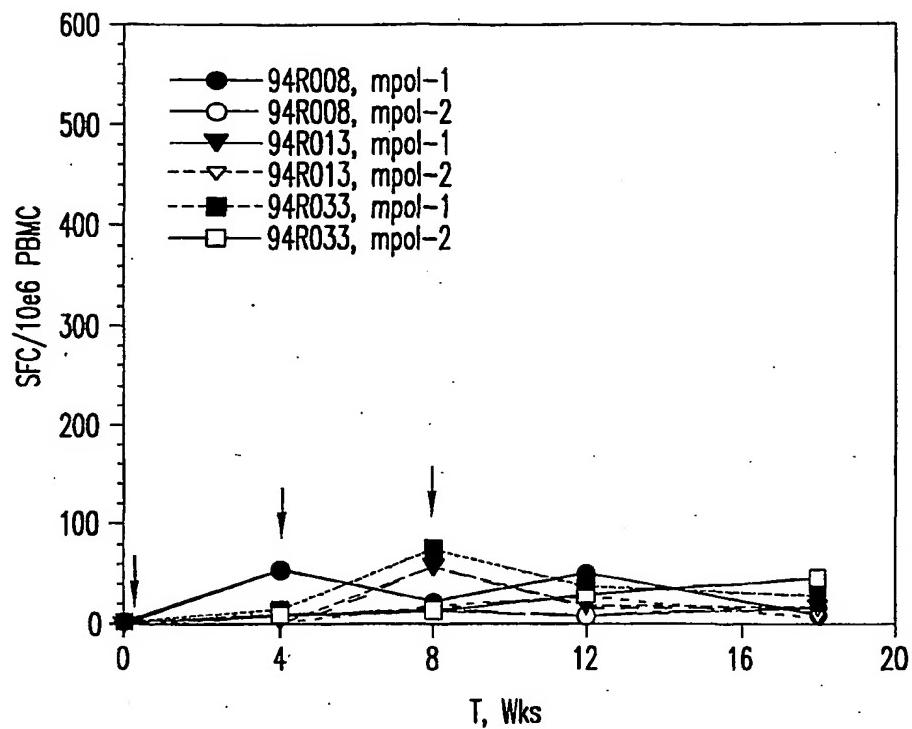


FIG.6A

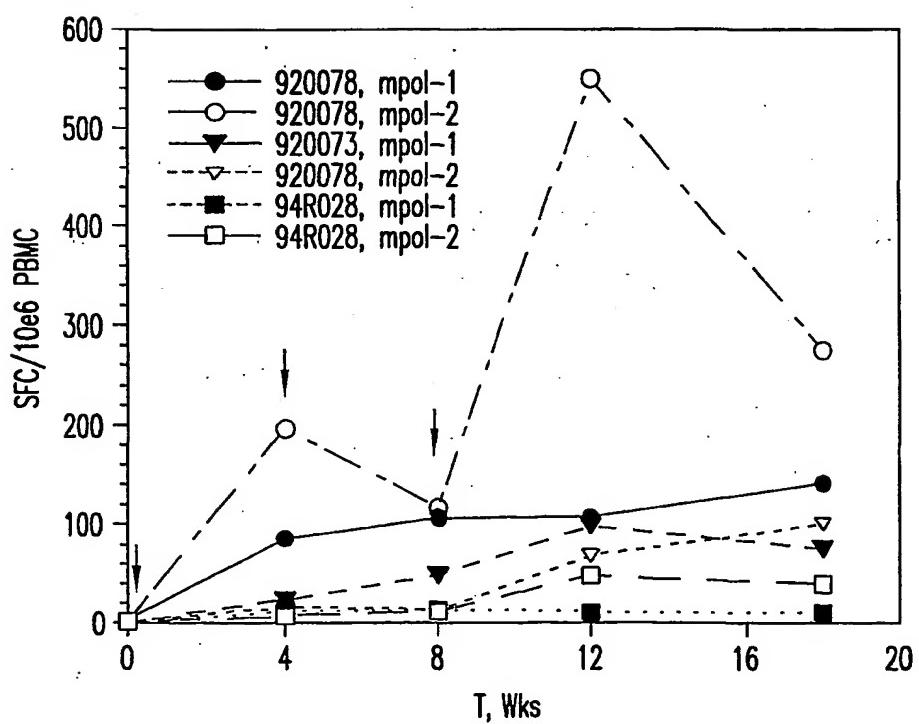


FIG.6B

8/11

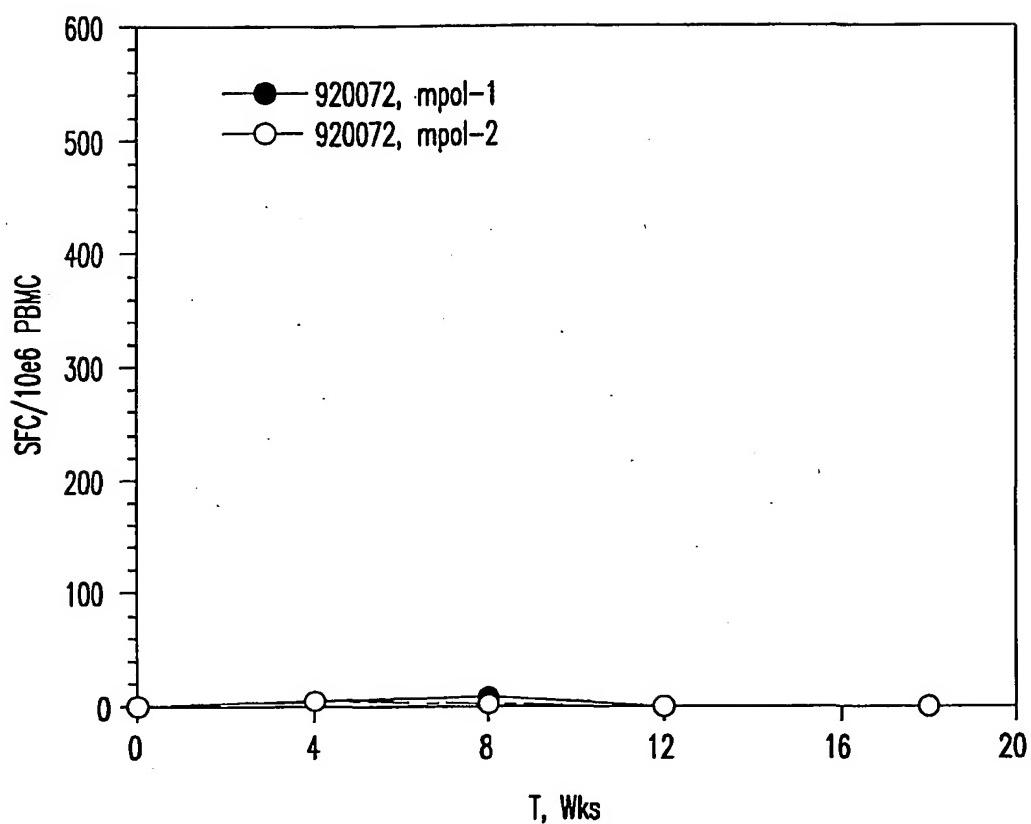


FIG.6C

9/11

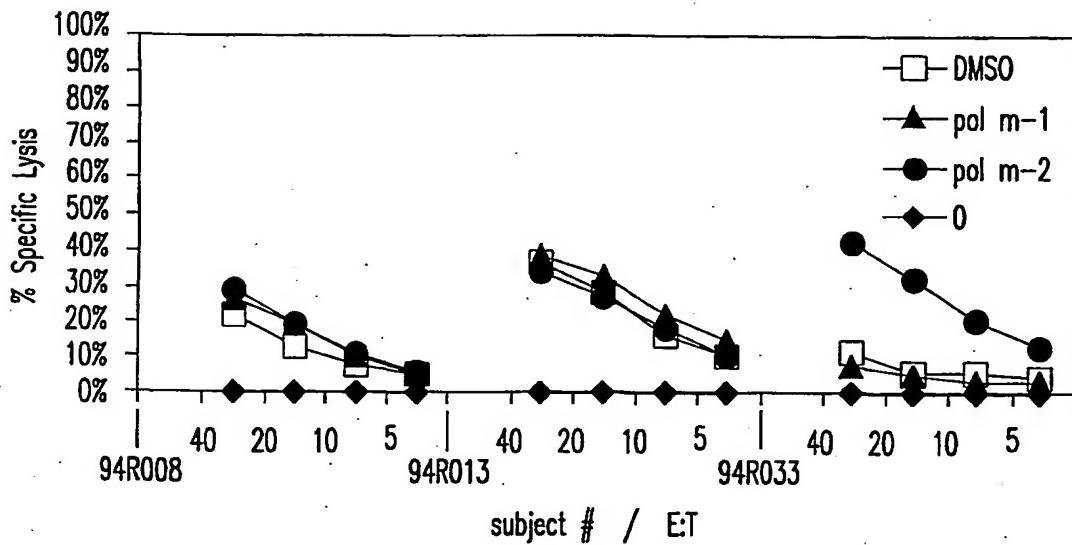


FIG.7A

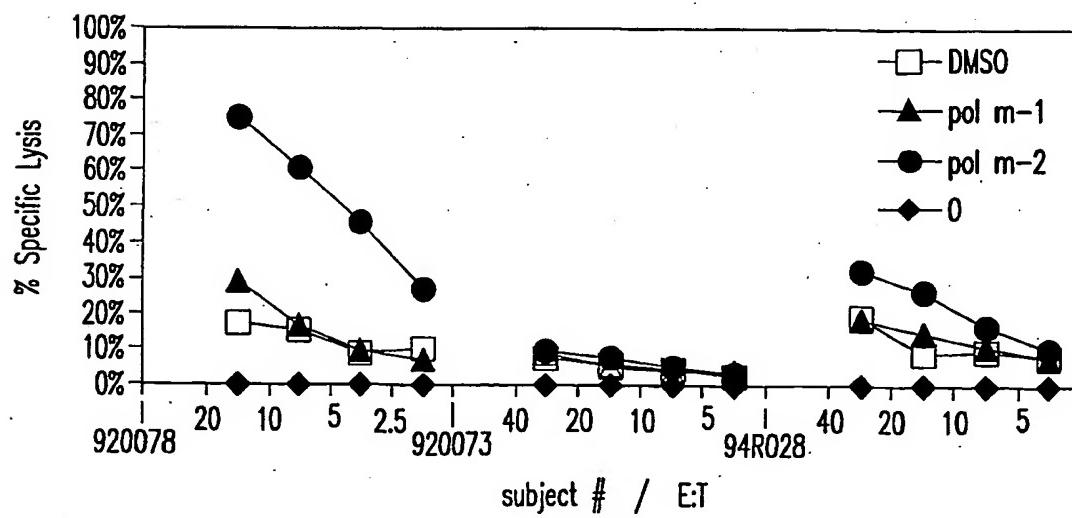


FIG.7B

10 / 11

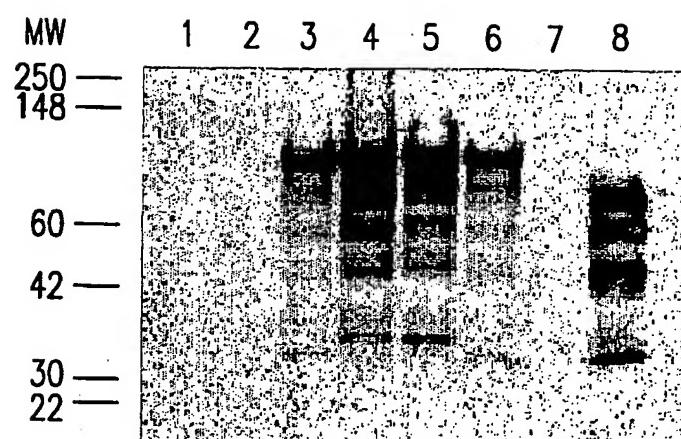


FIG.8

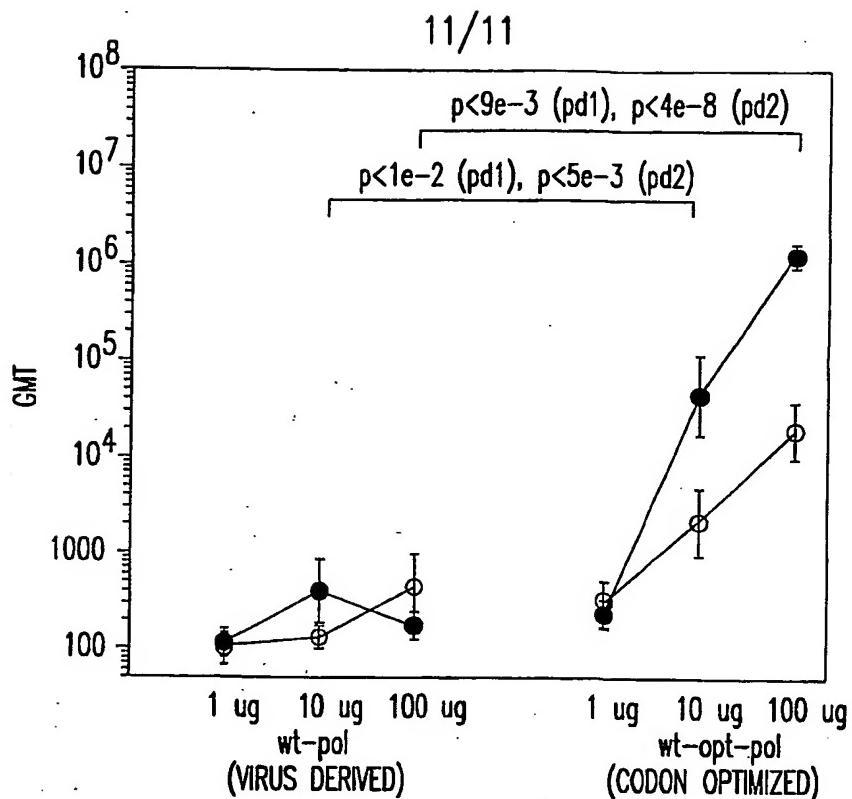


FIG.9

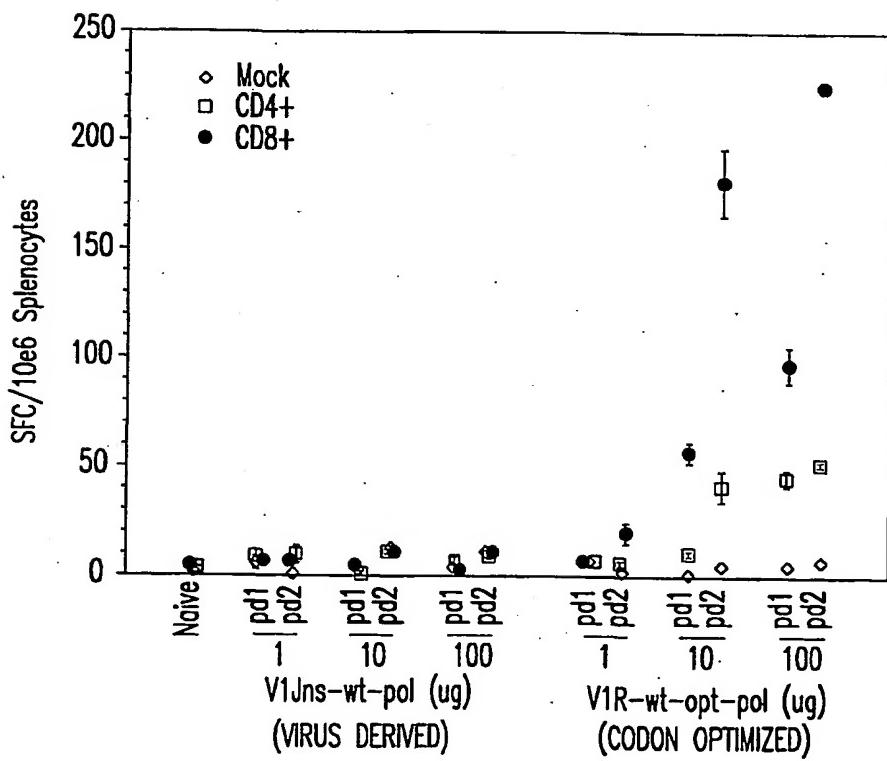


FIG.10

SEQUENCE LISTING

<110> Merck & Co., Inc.

<120> POLYNUCLEOTIDE VACCINES EXPRESSING CODON
OPTIMIZED HIV-1 POL AND MODIFIED HIV-1 POL

<130> 20608Y PCT

<160> 30

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2577

<212> DNA

<213> Human Immunodeficiency Virus-1

<220>

<221> CDS

<222> (10)...(2562)

<400> 1

agatctacc	atg	gcc	ccc	atc	tcc	ccc	att	gag	act	gtg	cct	gtg	aag	ctg		51
	Met	Ala	Pro	Ile	Ser	Pro	Ile	Glu	Thr	Val	Pro	Val	Lys	Leu		
	1				5				10							

aag	cct	ggc	atg	gat	ggc	ccc	aag	gtg	aag	cag	tgg	ccc	ctg	act	gag	99
Lys	Pro	Gly	Met	Asp	Gly	Pro	Lys	Val	Lys	Gln	Trp	Pro	Leu	Thr	Glu	
15				20					25				30			

gag	aag	atc	aag	gcc	ctg	gtg	gaa	atc	tgc	act	gag	atg	gag	aag	gag	147
Glu	Lys	Ile	Lys	Ala	Leu	Val	Glu	Ile	Cys	Thr	Glu	Met	Glu	Lys	Glu	
35					40				45							

ggc	aaa	atc	tcc	aag	att	ggc	ccc	gag	aac	ccc	tac	aac	acc	cct	gtg	195
Gly	Lys	Ile	Ser	Lys	Ile	Gly	Pro	Glu	Asn	Pro	Tyr	Asn	Thr	Pro	Val	
50					55				60							

ttt	gcc	atc	aag	aag	aag	gac	tcc	acc	aag	tgg	agg	aag	ctg	gtg	gac	243
Phe	Ala	Ile	Lys	Lys	Lys	Asp	Ser	Thr	Lys	Trp	Arg	Lys	Leu	Val	Asp	
65					70				75							

ttc	agg	gag	ctg	aac	aag	agg	acc	cag	gac	ttc	tgg	gag	gtg	cag	ctg	291
Phe	Arg	Glu	Leu	Asn	Lys	Arg	Thr	Gln	Asp	Phe	Trp	Glu	Val	Gln	Leu	
80					85				90							

ggc	atc	ccc	cac	ccc	gct	ggc	ctg	aag	aag	aag	tct	gtg	act	gtg		339
Gly	Ile	Pro	His	Pro	Ala	Gly	Leu	Lys	Lys	Lys	Ser	Val	Thr	Val		
95					100				105			110				

ctg	gat	gtg	ggg	gat	gcc	tac	ttc	tct	gtg	ccc	ctg	gat	gag	gac	ttc	387
Leu	Asp	Val	Gly	Asp	Ala	Tyr	Phe	Ser	Val	Pro	Leu	Asp	Glu	Asp	Phe	
115								120				125				

agg	aag	tac	act	gcc	tcc	acc	atc	ccc	tcc	atc	aac	aat	gag	acc	cct	435
Arg	Lys	Tyr	Thr	Ala	Phe	Thr	Ile	Pro	Ser	Ile	Asn	Asn	Glu	Thr	Pro	
130								135				140				

ggc atc agg tac cag tac aat gtg ctg ccc cag ggc tgg aag ggc tcc Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser 145 150 155	483
cct gcc atc ttc cag tcc tcc atg acc aag atc ctg gag ccc ttc agg Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg 160 165 170	531
aag cag aac cct gac att gtg atc tac cag tac atg gat gac ctg tat Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr 175 180 185 190	579
gtg ggc tct gac ctg gag att ggg cag cac agg acc aag att gag gag Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu 195 200 205	627
ctg agg cag cac ctg ctg agg tgg ggc ctg acc acc cct gac aag aag Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys 210 215 220	675
cac cag aag gag ccc ccc ttc ctg tgg atg ggc tat gag ctg cac ccc His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro 225 230 235	723
gac aag tgg act gtg cag ccc att gtg ctg cct gag aag gac tcc tgg Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp 240 245 250	771
act gtg aat gac atc cag aag ctg gtg ggc aag ctg aac tgg gcc tcc Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser 255 260 265 270	819
caa atc tac cct ggc atc aag gtg agg cag ctg tgc aag ctg ctg agg Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Arg 275 280 285	867
ggc acc aag gcc ctg act gag gtg atc ccc ctg act gag gag gct gag Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu 290 295 300	915
ctg gag ctg gct gag aac agg gag atc ctg aag gag cct gtg cat ggg Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly 305 310 315	963
gtg tac tat gac ccc tcc aag gac ctg att gct gag atc cag aag cag Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln 320 325 330	1011
ggc cag ggc cag tgg acc tac caa atc tac cag gag ccc ttc aag aac Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn 335 340 345 350	1059
ctg aag act ggc aag tat gcc agg atg agg ggg gcc cac acc aat gat Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp 355 360 365	1107
gtg aag cag ctg act gag gct gtg cag aag atc acc act gag tcc att Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile 370 375 380	1155

gtg atc tgg ggc aag acc ccc aag ttc aag ctg ccc atc cag aag gag Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu 385 390 395	1203
acc tgg gag acc tgg tgg act gag tac tgg cag gcc acc tgg atc cct Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro 400 405 410	1251
gag tgg gag ttt gtg aac acc ccc ccc ctg gtg aag ctg tgg tac cag Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln 415 420 425 430	1299
ctg gag aag gag ccc att gtg ggg gct gag acc ttc tat gtg gat ggg Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly 435 440 445	1347
gct gcc aac agg gag acc aag ctg ggc aag gct ggc tat gtg acc aac Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn 450 455 460	1395
agg ggc agg cag aag gtg gtg acc ctg act gac acc acc aac cag aag Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys 465 470 475	1443
act gag ctc cag gcc atc tac ctg gcc ctc cag gac tct ggc ctg gag Thr Glu Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu 480 485 490	1491
gtg aac att gtg act gac tcc cag tat gcc ctg ggc atc atc cag gcc Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala 495 500 505 510	1539
cag cct gat cag tct gag tct gag ctg gtg aac cag atc att gag cag Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln 515 520 525	1587
ctg atc aag aag gag aag gtg tac ctg gcc tgg gtg cct gcc cac aag Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys 530 535 540	1635
ggc att ggg ggc aat gag cag gtg gac aag ctg gtg tct gct ggc atc Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile 545 550 555	1683
agg aag gtg ctg ttc ctg gat ggc att gac aag gcc cag gat gag cat Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His 560 565 570	1731
gag aag tac cac tcc aac tgg agg gct atg gcc tct gac ttc aac ctg Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu 575 580 585 590	1779
ccc cct gtg gtg gct aag gag att gtg gcc tcc tgt gac aag tgc cag Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln 595 600 605	1827
ctg aag ggg gag gcc atg cat ggg cag gtg gac tgc tcc cct ggc atc Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile 610 615 620	1875

tgg cag ctg gac tgc acc cac ctg gag ggc aag gtg atc ctg gtg gct Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala 625 630 635	1923
gtg cat gtg gcc tcc ggc tac att gag gct gag gtg atc cct gct gag Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu 640 645 650	1971
aca ggc cag gag act gcc tac ttc ctg ctg aag ctg gct ggc agg tgg Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp 655 660 665 670	2019
cct gtg aag acc atc cac act gac aat ggc tcc aac ttc act ggg gcc Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala 675 680 685	2067
aca gtg agg gct gcc tgc tgg tgg gct ggc atc aag cag gag ttt ggc Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly 690 695 700	2115
atc ccc tac aac ccc cag tcc cag ggg gtg gtg gag tcc atg aac aag Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys 705 710 715	2163
gag ctg aag aag atc att ggg cag gtg agg gac cag gct gag cac ctg Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu 720 725 730	2211
aag aca gct gtg cag atg gct gtg ttc atc cac aac ttc aag agg aag Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys 735 740 745 750	2259
ggg ggc atc ggg ggc tac tcc gct ggg gag agg att gtg gac atc att Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile 755 760 765	2307
gcc aca gac atc cag acc aag gag ctc cag aag cag atc acc aag atc Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile 770 775 780	2355
cag aac ttc agg gtg tac tac agg gac tcc agg aac ccc ctg tgg aag Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys 785 790 795	2403
ggc cct gcc aag ctg ctg tgg aag ggg gag ggg gct gtg gtg atc cag Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln 800 805 810	2451
gac aac tct gac atc aag gtg gtg ccc agg agg aag gcc aag atc atc Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile 815 820 825 830	2499
agg gac tat ggc aag cag atg gct ggg gat gac tgt gtg gcc tcc agg Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg 835 840 845	2547
cag gat gag gac taa agccgggca gatct Gln Asp Glu Asp * 850	2577

<211> 850

<212> PRT

<213> Human Immunodeficiency Virus-1

<400> 2

Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro
 1 5 10 15
 Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys
 20 25 30
 Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys
 35 40 45
 Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala
 50 55 60
 Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg
 65 70 75 80
 Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile
 85 90 95
 Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Asp
 100 105 110
 Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys
 115 120 125
 Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile
 130 135 140
 Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
 145 150 155 160
 Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Pro Phe Arg Lys Gln
 165 170 175
 Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly
 180 185 190
 Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
 195 200 205
 Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln
 210 215 220
 Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
 225 230 235 240
 Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
 245 250 255
 Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
 260 265 270
 Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr
 275 280 285
 Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
 290 295 300
 Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
 305 310 315 320
 Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
 325 330 335
 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
 340 345 350
 Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
 355 360 365
 Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
 370 375 380
 Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
 385 390 395 400
 Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
 405 410 415
 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
 420 425 430
 Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala
 435 440 445

Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
 450 455 460
 Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu
 465 470 475 480
 Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
 485 490 495
 Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
 500 505 510
 Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
 515 520 525
 Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
 530 535 540
 Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys
 545 550 555 560
 Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
 565 570 575
 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
 580 585 590
 Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
 595 600 605
 Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
 610 615 620
 Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
 625 630 635 640
 Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
 645 650 655
 Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
 660 665 670
 Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
 675 680 685
 Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
 690 695 700
 Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu
 705 710 715 720
 Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
 725 730 735
 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
 740 745 750
 Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
 755 760 765
 Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
 770 775 780
 Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
 785 790 795 800
 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
 805 810 815
 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
 820 825 830
 Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
 835 840 845
 Glu Asp
 850

<210> 3
 <211> 2577
 <212> DNA
 <213> Human Immunodeficiency Virus-1

 <220>
 <221> CDS
 <222> (10)...(2562).

<400> 3
 agatctacc atg gcc ccc atc tcc ccc att gag act gtg cct gtg aag ctg
 Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu
 1 5 10

 aag cct ggc atg gat ggc ccc aag gtg aag cag tgg ccc ctg act gag
 Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu
 15 20 25 30

 gag aag atc aag gcc ctg gtg gaa atc tgc act gag atg gag aag gag
 Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu
 35 40 45 50

 ggc aaa atc tcc aag att ggc ccc gag aac ccc tac aac acc cct gtg
 Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val
 50 55 60 65

 ttt gcc atc aag aag aag gac tcc acc aag tgg agg aag ctg gtg gac
 Phe Ala Ile Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp
 65 70 75 80

 ttc agg gag ctg aac aag agg acc cag gac ttc tgg gag gtg cag ctg
 Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu
 80 85 90 95

 ggc atc ccc cac ccc gct ggc ctg aag aag aag aag tct gtg act gtg
 Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val
 95 100 105 110

 ctg gct gtg ggg gat gcc tac ttc tct gtg ccc ctg gat gag gac ttc
 Leu Ala Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe
 115 120 125 130

 agg aag tac act gcc ttc acc atc ccc tcc atc aac aat gag acc cct
 Arg Lys Tyr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro
 130 135 140 145

 ggc atc agg tac cag tac aat gtg ctg ccc cag ggc tgg aag ggc tcc
 Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser
 145 150 155 160

 cct gcc atc ttc cag tcc tcc atg acc aag atc ctg gag ccc ttc agg
 Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg
 160 165 170 175

 aag cag aac cct gac att gtg atc tac cag tac atg gct gcc ctg tat
 Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr
 175 180 185 190

 gtg ggc tct gac ctg gag att ggg cag cac agg acc aag att gag gag
 Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu
 195 200 205 210

 ctg agg cag cac ctg ctg agg tgg ggc ctg acc acc cct gac aag aag
 Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys
 210 215 220 225

 cac cag aag gag ccc ccc ttc ctg tgg atg ggc tat gag ctg cac ccc
 His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro
 225 230 235 240

gac aag tgg act gtg cag ccc att gtg ctg cct gag aag gac tcc tgg Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp 240 245 250	771
act gtg aat gac atc cag aag ctg gtg ggc aag ctg aac tgg gcc tcc Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser 255 260 265 270	819
caa atc tac cct ggc atc aag gtg agg cag ctg tgc aag ctg ctg agg Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Ile Arg 275 280 285	867
ggc acc aag gcc ctg act gag gtg atc ccc ctg act gag gag gct gag Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu 290 295 300	915
ctg gag ctg gct gag aac agg gag atc ctg aag gag cct gtg cat ggg Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly 305 310 315	963
gtg tac tat gac ccc tcc aag gac ctg att gct gag atc cag aag gag Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln 320 325 330	1011
ggc cag ggc cag tgg acc tac caa atc tac cag gag ccc ttc aag aac Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn 335 340 345 350	1059
ctg aag act ggc aag tat gcc agg atg agg ggg gcc cac acc aat gat Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp 355 360 365	1107
gtg aag cag ctg act gag gct gtg cag aag atc acc act gag tcc att Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile 370 375 380	1155
gtg atc tgg ggc aag acc ccc aag ttc aag ctg ccc atc cag aag gag Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu 385 390 395	1203
acc tgg gag acc tgg tgg act gag tac tgg cag gcc acc tgg atc cct Thr Trp Glu Thr Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro 400 405 410	1251
gag tgg gag ttt gtg aac acc ccc ccc ctg gtg aag ctg tgg tac cag Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln 415 420 425 430	1299
ctg gag aag gag ccc att gtg ggg gct gag acc ttc tat gtg gct ggg Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly 435 440 445	1347
gct gcc aac agg gag acc aag ctg ggc aag gct ggc tat gtg acc aac Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn 450 455 460	1395
agg ggc agg cag aag gtg gtg acc ctg act gac acc acc aac cag aag Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys 465 470 475	1443

act gcc ctc cag gcc atc tac ctg gcc ctc cag gac tct ggc ctg gag		1491
Thr Ala Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu		
480 485 490		
gtg aac att gtg act gcc tcc cag tat gcc ctg ggc atc atc cag gcc		1539
Val Asn Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala		
495 500 505 510		
cag cct gat cag tct gag tct gag ctg gtg aac cag atc att gag cag		1587
Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln		
515 520 525		
ctg atc aag aag gag aag gtg tac ctg gcc tgg gtg cct gcc cac aag		1635
Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys		
530 535 540		
ggc att ggg ggc aat gag cag gtg gac aag ctg gtg tct gct ggc atc		1683
Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile		
545 550 555		
agg aag gtg ctg ttc ctg gat ggc att gac aag gcc cag gat gag cat		1731
Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His		
560 565 570		
gag aag tac cac tcc aac tgg agg gct atg gcc tct gac ttc aac ctg		1779
Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu		
575 580 585 590		
ccc cct gtg gtg gct aag gag att gtg gcc tcc tgt gac aag tgc cag		1827
Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln		
595 600 605		
ctg aag ggg gag gcc atg cat ggg cag gtg gac tgc tcc cct ggc atc		1875
Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile		
610 615 620		
tgg cag ctg gcc tgc acc cac ctg gag ggc aag gtg atc ctg gtg gct		1923
Trp Gln Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala		
625 630 635		
gtg cat gtg gcc tcc ggc tac att gag gct gag gtg atc cct gct gag		1971
Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu		
640 645 650		
aca ggc cag gag act gcc tac ttc ctg ctg aag ctg gct ggc agg tgg		2019
Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp		
655 660 665 670		
cct gtg aag acc atc cac act gcc aat ggc tcc aac ttc act ggg gcc		2067
Pro Val Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala		
675 680 685		
aca gtg agg gct gcc tgc tgg tgg gct ggc atc aag cag gag ttt ggc		2115
Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly		
690 695 700		
atc ccc tac aac ccc cag tcc cag ggg gtg gtg gcc tcc atg aac aag		2163
Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys		
705 710 715		

gag ctg aag aag atc att ggg cag gtg agg gac cag gct gag cac ctg Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu 720 725 730	2211
aag aca gct gtg cag atg gct gtg ttc atc cac aac ttc aag agg aag Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys 735 740 745 750	2259
ggg ggc atc ggg ggc tac tcc gct ggg gag agg att gtg gac atc att Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile 755 760 765	2307
gcc aca gac atc cag acc aag gag ctc cag aag cag atc acc aag atc Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile 770 775 780	2355
cag aac ttc agg gtg tac tac agg gac tcc agg aac ccc ctg tgg aag Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys 785 790 795	2403
ggc cct gcc aag ctg ctg tgg aag ggg gag ggg gct gtg gtg atc cag Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln 800 805 810	2451
gac aac tct gac atc aag gtg gtg ccc agg agg aag gcc aag atc atc Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile 815 820 825 830	2499
agg gac tat ggc aag cag atg gct ggg gat gac tgt gtg gcc tcc agg Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg 835 840 845	2547
cag gat gag gac taa agccggca gatct Gln Asp Glu Asp * 850	2577

<210> 4
<211> 850
<212> PRT
<213> Human Immunodeficiency Virus-1

<400> 4	
Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro	
1 5 10 15	
Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys	
20 25 30	
Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys	
35 40 45	
Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala	
50 55 60	
Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg	
65 70 75 80	
Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile	
85 90 95	
Pro His Pro Ala Gly Leu Lys Lys Ser Val Thr Val Leu Ala	
100 105 110	
Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys	
115 120 125	
Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile	
130 135 140	

Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala
 145 150 155 160
 Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
 165 170 175
 Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly
 180 185 190
 Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg
 195 200 205
 Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln
 210 215 220
 Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys
 225 230 235 240
 Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val
 245 250 255
 Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile
 260 265 270
 Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Arg Gly Thr
 275 280 285
 Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu
 290 295 300
 Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr
 305 310 315 320
 Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln
 325 330 335
 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys
 340 345 350
 Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys
 355 360 365
 Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile
 370 375 380
 Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp
 385 390 395 400
 Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
 405 410 415
 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu
 420 425 430
 Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala
 435 440 445
 Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
 450 455 460
 Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala
 465 470 475 480
 Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
 485 490 495
 Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
 500 505 510
 Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
 515 520 525
 Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
 530 535 540
 Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys
 545 550 555 560
 Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
 565 570 575
 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
 580 585 590
 Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
 595 600 605
 Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
 610 615 620
 Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
 625 630 635 640

<210> 5
<211> 2650
<212> DNA

<212> DNA
<213> Human

<213> Human Immunodeficiency Virus-1

220

221

<221> CDS
<222> (8)

<222> 10) ... (2055)

400 5

gatcacc

```

gateacc tcy gac gca atg aag aya ggg ccc tgc ctc gtc gtc gtc ctc ctc ctc
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu
      1           5           10

```

49

```

tgt gga gca gtc ttc gtt tcg ccc agc gag atc tcc gcc ccc atc tcc
Cys Gly Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser
   15           20           25           30

```

97

ccc att gag act gtg cct gtg aag ctg aag cct ggc atg gat ggc ccc
 Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro
 35 40 45

145

```

aag gtg aag cag tgg ccc ctg act gag gag aag atc aag gcc ctg gtg
Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val
      50           55           .   60

```

193

gaa atc tgc act gag atg gag aag gag ggc aaa atc tcc aag att ggc
 Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly
 65 70 75

241

ccc gag aac ccc tac aac acc cct gtg ttt gcc atc aag aag aag gac
Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp
80 85 90

280

tcc acc aag tgg agg aag ctg gtg gac ttc agg gag ctg aac aag agg		337
Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg		
95 100 105 110		
acc cag gac ttc tgg gag gtg cag ctg ggc atc ccc cac ccc gct ggc		385
Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly		
115 120 125		
ctg aag aag aag aag tct gtg act gtg ctg gat gtg ggg gat gcc tac		433
Leu Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr		
130 135 140		
ttc tct gtg ccc ctg gat gag gac ttc agg aag tac act gcc ttc acc		481
Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr		
145 150 155		
atc ccc tcc atc aac aat gag acc cct ggc atc agg tac cag tac aat		529
Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn		
160 165 170		
gtg ctg ccc cag ggc tgg aag ggc tcc cct gcc atc ttc cag tcc tcc		577
Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser		
175 180 185 190		
atg acc aag atc ctg gag ccc ttc agg aag cag aac cct gac att gtg		625
Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val		
195 200 205		
atc tac cag tac atg gat gac ctg tat gtg ggc tct gac ctg gag att		673
Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile		
210 215 220		
ggg cag cac agg acc aag att gag gag ctg agg cag cac ctg ctg agg		721
Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg		
225 230 235		
tgg ggc ctg acc acc cct gac aag aag cac cag aag gag ccc ccc ttc		769
Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe		
240 245 250		
ctg tgg atg ggc tat gag ctg cac ccc gac aag tgg act gtg cag ccc		817
Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro		
255 260 265 270		
att gtg ctg cct gag aag gac tcc tgg act gtg aat gac atc cag aag		865
Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys		
275 280 285		
ctg gtg ggc aag ctg aac tgg gcc tcc caa atc tac cct ggc atc aag		913
Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys		
290 295 300		
gtg agg cag ctg tgc aag ctg ctg agg ggc acc aag gcc ctg act gag		961
Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu		
305 310 315		
gtg atc ccc ctg act gag gag gct gag ctg gag ctg gct gag aac agg		1009
Val Ile Pro Leu Thr Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg		
320 325 330		

gag atc ctg aag gag cct gtg cat ggg gtg tac tat gac ccc tcc aag Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys 335 340 345 350	1057
gac ctg att gct gag atc cag aag cag ggc cag ggc cag tgg acc tac Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr 355 360 365	1105
caa atc tac cag gag ccc ttc aag aac ctg aag act ggc aag tat gcc Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala 370 375 380	1153
agg atg agg ggg gcc cac acc aat gat gtg aag cag ctg act gag gct Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala 385 390 395	1201
gtg cag aag atc acc act gag tcc att gtg atc tgg ggc aag acc ccc Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro 400 405 410	1249
aag ttc aag ctg ccc atc cag aag gag acc tgg gag acc tgg tgg act Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr 415 420 425 430	1297
gag tac tgg cag gcc acc tgg atc cct gag tgg gag ttt gtg aac acc Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr 435 440 445	1345
ccc ccc ctg gtg aag ctg tgg tac cag ctg gag aag gag ccc att gtg Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val 450 455 460	1393
ggg gct gag acc ttc tat gtg gat ggg gct gcc aac agg gag acc aag Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys 465 470 475	1441
ctg ggc aag gct ggc tat gtg acc aac agg ggc agg cag aag gtg gtg Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val 480 485 490	1489
acc ctg act gac acc acc aac cag aag act gag ctc cag gcc atc tac Thr Leu Thr Asp Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr 495 500 505 510	1537
ctg gcc ctc cag gac tct ggc ctg gag gtg aac att gtg act gac tcc Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser 515 520 525	1585
cag tat gcc ctg ggc atc atc cag gcc cag cct gat cag tct gag tct Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser 530 535 540	1633
gag ctg gtg aac cag atc att gag cag ctg atc aag aag gag aag gtg Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val 545 550 555	1681
tac ctg gcc tgg gtg cct gcc ctc aag ggc att ggg ggc aat gag cag Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln 560 565 570	1729

gtg gac aag ctg gtg tct gct ggc atc agg aag gtg ctg ttc ctg gat Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp 575 580 585 590	1777
ggc att gac aag gcc cag gat gag cat gag aag tac cac tcc aac tgg Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp 595 600 605	1825
agg gct atg gcc tct gac ttc aac ctg ccc cct gtg gtg gct aag gag Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu 610 615 620	1873
att gtg gcc tcc tgt gac aag tgc cag ctg aag ggg gag gcc atg cat Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His 625 630 635	1921
ggg cag gtg gac tgc tcc cct ggc atc tgg cag ctg gac tgc acc cac Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His 640 645 650	1969
ctg gag ggc aag gtg atc ctg gtg gct gtg cat gtg gcc tcc ggc tac Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr 655 660 665 670	2017
att gag gct gag gtg atc cct gct gag aca ggc cag gag act gcc tac Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr 675 680 685	2065
ttc ctg ctg aag ctg gct ggc agg tgg cct gtg aag acc atc cac act Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr 690 695 700	2113
gac aat ggc tcc aac ttc act ggg gcc aca gtg agg gct gcc tgc tgg Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp 705 710 715	2161
tgg gct ggc atc aag cag gag ttt ggc atc ccc tac aac ccc cag tcc Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser 720 725 730	2209
cag ggg gtg gtg gag tcc atg aac aag gag ctg aag aag atc att ggg Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu Lys Ile Ile Gly 735 740 745 750	2257
cag gtg agg gac cag gct gag cac ctg aag aca gct gtg cag atg gct Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala 755 760 765	2305
gtg ttc atc cac aac ttc aag agg aag ggg ggc atc ggg ggc tac tcc Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser 770 775 780	2353
gct ggg gag agg att gtg gac atc att gcc aca gac atc cag acc aag Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys 785 790 795	2401
gag ctc cag aag cag atc acc aag atc cag aac ttc agg gtg tac tac Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr 800 805 810	2449

agg gac tcc agg aac ccc ctg tgg aag ggc cct gcc aag ctg ctg tgg 2497
 Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp
 815 820 825 830
 aag ggg gag ggg gct gtg gtg atc cag gac aac tct gac atc aag gtg 2545
 Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val
 835 840 845
 gtg ccc agg agg aag gcc aag atc atc agg gac tat ggc aag cag atg 2593
 Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met
 850 855 860
 gct ggg gat gac tgt gtg gcc tcc agg cag gat gag gac taa 2635
 Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp *
 865 870 875
 agccccggca gatct 2650
 <210> 6
 <211> 875
 <212> PRT
 <213> Human Immunodeficiency Virus-1
 <400> 6
 Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
 1 5 10 15
 Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
 20 25 30
 Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
 35 40 45
 Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
 50 55 60
 Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
 65 70 75 80
 Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
 85 90 95
 Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
 100 105 110
 Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
 115 120 125
 Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser
 130 135 140
 Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
 145 150 155 160
 Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
 165 170 175
 Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
 180 185 190
 Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
 195 200 205
 Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
 210 215 220
 His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
 225 230 235 240
 Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
 245 250 255
 Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
 260 265 270
 Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
 275 280 285
 Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
 290 295 300

Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
 305 310 315 320
 Pro Leu Thr Glu Glu Ala Glu Leu Glu Ala Glu Asn Arg Glu Ile
 325 330 335
 Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
 340 345 350
 Ile Ala Glu Ile Gln Lys Gln Gly Gln Trp Thr Tyr Gln Ile
 355 360 365
 Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
 370 375 380
 Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln
 385 390 395 400
 Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
 405 410 415
 Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Thr Glu Tyr
 420 425 430
 Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
 435 440 445
 Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
 450 455 460
 Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
 465 470 475 480
 Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
 485 490 495
 Thr Asp Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala
 500 505 510
 Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr
 515 520 525
 Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
 530 535 540
 Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
 545 550 555 560
 Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
 565 570 575
 Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
 580 585 590
 Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
 595 600 605
 Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
 610 615 620
 Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
 625 630 635 640
 Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu
 645 650 655
 Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
 660 665 670
 Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
 675 680 685
 Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn
 690 695 700
 Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala
 705 710 715 720
 Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
 725 730 735
 Val Val Glu Ser Met Asn Lys Glu Leu Lys Ile Ile Gly Gln Val
 740 745 750
 Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
 755 760 765
 Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly
 770 775 780
 Glu Arg Ile Val Asp Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
 785 790 795 800

Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
 805 810 815
 Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
 820 825 830
 Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
 835 840 845
 Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
 850 855 860
 Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
 865 870 875

<210> 7

<211> 2650

<212> DNA

<213> Human Immunodeficiency Virus-1

<220>

<221> CDS

<222> (8) ... (2635)

<400> 7

gatcacc	atg	gat	gca	atg	aag	aga	ggg	ctc	tgc	tgt	gtg	ctg	ctg	ctg		49
Met	Asp	Ala	Met	Lys	Arg	Gly	Leu	Cys	Cys	Val	Leu	Leu	Leu			
1			5							10						

tgt	gga	gca	gtc	ttc	gtt	tgc	ccc	agc	gag	atc	tcc	gcc	ccc	atc	tcc	97
Cys	Gly	Ala	Val	Phe	Val	Ser	Pro	Ser	Glu	Ile	Ser	Ala	Pro	Ile	Ser	
15				20.						25			30			

ccc	att	gag	act	gtg	cct	gtg	aag	ctg	aag	cct	ggc	atg	gat	ggc	ccc	145
Pro	Ile	Glu	Thr	Val	Pro	Val	Lys	Leu	Lys	Pro	Gly	Met	Asp	Gly	Pro	
35				40								45				

aag	gtg	aag	cag	tgg	ccc	ctg	act	gag	gag	aag	atc	aag	gcc	ctg	gtg	193
Lys	Val	Lys	Gln	Trp	Pro	Leu	Thr	Glu	Glu	Lys	Ile	Lys	Ala	Leu	Val	
50				55							60					

gaa	atc	tgc	act	gag	atg	gag	aag	ggc	aaa	atc	tcc	aag	att	ggc	241
Glu	Ile	Cys	Thr	Glu	Met	Glu	Lys	Glu	Gly	Ile	Ser	Lys	Ile	Gly	
65				70							75				

ccc	gag	aac	ccc	tac	aac	acc	cct	gtg	ttt	gcc	atc	aag	aag	gac	289
Pro	Glu	Asn	Pro	Tyr	Asn	Thr	Pro	Val	Phe	Ala	Ile	Lys	Lys	Asp	
80				85						90					

tcc	acc	aag	tgg	agg	aag	ctg	gtg	gac	ttc	agg	gag	ctg	aac	aag	agg	337
Ser	Thr	Lys	Trp	Arg	Lys	Leu	Val	Asp	Phe	Arg	Glu	Leu	Asn	Lys	Arg	
95				100						105			110			

acc	cag	gac	ttc	tgg	gag	gtg	cag	ctg	ggc	atc	ccc	cac	ccc	gct	ggc	385
Thr	Gln	Asp	Phe	Trp	Glu	Val	Gln	Leu	Gly	Ile	Pro	His	Pro	Ala	Gly	
115				120							125					

ctg	aag	aag	aag	tct	gtg	act	gtg	ctg	gct	gtg	ggg	gat	gcc	tac	433
Leu	Lys	Lys	Lys	Ser	Val	Thr	Val	Leu	Ala	Val	Gly	Asp	Ala	Tyr	
130				135						140					

ttc	tct	gtg	ccc	ctg	gat	gag	gac	ttc	agg	aag	tac	act	gcc	ttc	acc	481
Phe	Ser	Val	Pro	Leu	Asp	Glu	Asp	Phe	Arg	Lys	Tyr	Thr	Ala	Phe	Thr	
145				150						155						

atc ccc tcc atc aac aat gag acc cct ggc atc agg tac cag tac aat Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn 160 165 170	529
gtg ctg ccc cag ggc tgg aag ggc tcc cct gcc atc ttc cag tcc tcc Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser 175 180 185 190	577
atg acc aag atc ctg gag ccc ttc agg aag cag aac cct gac att gtg Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val 195 200 205	625
atc tac cag tac atg gct gcc ctg tat gtg ggc tct gac ctg gag att Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile 210 215 220	673
ggg cag cac agg acc aag att gag gag ctg agg cag cac ctg ctg agg Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg 225 230 235	721
tgg ggc ctg acc acc cct gac aag aag cac cag aag gag ccc ccc ttc Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe 240 245 250	769
ctg tgg atg ggc tat gag ctg cac ccc gac aag tgg act gtg cag ccc Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro 255 260 265 270	817
att gtg ctg cct gag aag gac tcc tgg act gtg aat gac atc cag aag Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys 275 280 285	865
ctg gtg ggc aag ctg aac tgg gcc tcc caa atc tac cct ggc atc aag Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys 290 295 300	913
gtg agg cag ctg tgc aag ctg ctg agg ggc acc aag gcc ctg act gag Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu 305 310 315	961
gtg atc ccc ctg act gag gag gct gag ctg gag ctg gct gag aac agg Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg 320 325 330	1009
gag atc ctg aag gag cct gtg cat ggg gtg tac tat gac ccc tcc aag Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys 335 340 345 350	1057
gac ctg att gct gag atc cag aag cag ggc cag ggc cag tgg acc tac Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gln Trp Thr Tyr 355 360 365	1105
caa atc tac cag gag ccc ttc aag aac ctg aag act ggc aag tat gcc Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala 370 375 380	1153
agg atg agg ggg gcc cac acc aat gat gtg aag cag ctg act gag gct Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala 385 390 395	1201

gtg cag aag atc acc act gag tcc att gtg atc tgg ggc aag acc ccc Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro 400 405 410	1249
aag ttc aag ctg ccc atc cag aag gag acc tgg gag acc tgg tgg act Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr 415 420 425 430	1297
gag tac tgg cag gcc acc tgg atc cct gag tgg gag ttt gtg aac acc Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr 435 440 445	1345
ccc ccc ctg gtg aag ctg tgg tac cag ctg gag aag gag ccc att gtg Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val 450 455 460	1393
ggg gct gag acc ttc tat gtg gct ggg gct gcc aac agg gag acc aag Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys 465 470 475	1441
ctg ggc aag gct ggc tat gtg acc aac agg ggc agg cag aag gtg gtg Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val 480 485 490	1489
acc ctg act gac acc acc aac cag aag act gcc ctc cag gcc atc tac Thr Leu Thr Asp Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr 495 500 505 510	1537
ctg gcc ctc cag gac tct ggc ctg gag gtg aac att gtg act gcc tcc Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser 515 520 525	1585
cag tat gcc ctg ggc atc atc cag gcc cag cct gat cag tct gag tct Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser 530 535 540	1633
gag ctg gtg aac cag atc att gag cag ctg atc aag aag gag aag gtg Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val 545 550 555	1681
tac ctg gcc tgg gtg cct gcc cac aag ggc att ggg ggc aat gag cag Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln 560 565 570	1729
gtg gac aag ctg gtg tct gct ggc atc agg aag gtg ctg ttc ctg gat Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp 575 580 585 590	1777
ggc att gac aag gcc cag gat gag cat gag aag tac cac tcc aac tgg Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp 595 600 605	1825
agg gct atg gcc tct gac ttc aac ctg ccc cct gtg gtg gct aag gag Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu 610 615 620	1873
att gtg gcc tcc tgt gac aag tgc cag ctg aag ggg gag gcc atg cat Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His 625 630 635	1921

ggg cag gtg gac tgc tcc cct ggc atc tgg cag ctg gcc tgc acc cac Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His 640 645 650	1969
ctg gag ggc aag gtg atc ctg gtg gct gtg cat gtg gcc tcc ggc tac Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr 655 660 665 670	2017
att gag gct gag gtg atc cct gct gag aca ggc cag gag act gcc tac Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr 675 680 685	2065
tcc ctg ctg aag ctg gct ggc agg tgg cct gtg aag acc atc cac act Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr 690 695 700	2113
gcc aat ggc tcc aac ttc act ggg gcc aca gtc agg gct gcc tgc tgg Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp 705 710 715	2161
tgg gct ggc atc aag cag gag ttt ggc atc ccc tac aac ccc cag tcc Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser 720 725 730	2209
cag ggg gtg gtg gcc tcc atg aac aag gag ctg aag aag atc att ggg Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly 735 740 745 750	2257
cag gtg agg gac cag gct gag cac ctg aag aca gct gtg cag atg gct Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala 755 760 765	2305
gtg ttc atc cac aac ttc aag agg aag ggg ggc atc ggg ggc tac tcc Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser 770 775 780	2353
gct ggg gag agg att gtg gac atc att gcc aca gac atc cag acc aag Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys 785 790 795	2401
gag ctc cag aag cag atc acc aag atc cag aac ttc agg gtg tac tac Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr 800 805 810	2449
agg gac tcc agg aac ccc ctg tgg aag ggc cct gcc aag ctg ctg tgg Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp 815 820 825 830	2497
aag ggg gag ggg gct gtg gtg atc cag gac aac tct gac atc aag gtg Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val 835 840 845	2545
gtg ccc agg agg aag gcc aag atc atc agg gac tat ggc aag cag atg Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met 850 855 860	2593
gct ggg gat gac tgt gtg gcc tcc agg cag gat gag gac taa Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp *865 870 875	2635
agccccgggca gatct	2650

<210> 8
<211> 875
<212> PRT

<213> Human Immunodeficiency Virus-1

<400> 8

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
1 5 10 15
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
20 25 30
Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
35 40 45
Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
50 55 60
Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
65 70 75 80
Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
85 90 95
Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
100 105 110
Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
115 120 125
Lys Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser
130 135 140
Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
145 150 155 160
Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
165 170 175
Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
180 185 190
Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
195 200 205
Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
210 215 220
His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
225 230 235 240
Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
245 250 255
Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
260 265 270
Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
275 280 285
Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
290 295 300
Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
305 310 315 320
Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
325 330 335
Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
340 345 350
Ile Ala Glu Ile Gln Lys Gln Gly Gln Gln Trp Thr Tyr Gln Ile
355 360 365
Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
370 375 380
Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln
385 390 395 400
Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
405 410 415
Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr
420 425 430
Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
435 440 445

Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
 450 455 460
 Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
 465 470 475 480
 Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
 485 490 495
 Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala
 500 505 510
 Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr
 515 520 525
 Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
 530 535 540
 Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
 545 550 555 560
 Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
 565 570 575
 Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
 580 585 590
 Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
 595 600 605
 Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
 610 615 620
 Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
 625 630 635 640
 Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu
 645 650 655
 Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
 660 665 670
 Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
 675 680 685
 Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn
 690 695 700
 Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala
 705 710 715 720
 Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
 725 730 735
 Val Val Ala Ser Met Asn Lys Glu Leu Lys Ile Ile Gly Gln Val
 740 745 750
 Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
 755 760 765
 Ile His Asn Phe Lys Arg Lys Gly Ile Gly Gly Tyr Ser Ala Gly
 770 775 780
 Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
 785 790 795 800
 Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
 805 810 815
 Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
 820 825 830
 Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
 835 840 845
 Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
 850 855 860
 Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
 865 870 875

<210> 9
 <211> 4945
 <212> DNA
 <213> E. coli (V1Jns-tpa)

<400> 9
 tcgcgcgttt cggatgac ggtgaaaacc tctgacacat gcagctcccg gagacggta

cagcttgtct	gtaagcggat	gccgggagca	gacaaggccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatacaga	gcagattgt	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgttaaggag	aaaataccgc	atcagattgg	240
ctattggca	ttgcatacgt	tgtatccata	tcataaatatg	tacatttata	ttggctcatg	300
tccaaacatta	ccgccccatgtt	gacattgatt	attgactagt	tattaatagt	aatcaattac	360
ggggtcatt	gttcatacgcc	cataatatgga	gttccgcgtt	acataactta	cggtaaatgg	420
ccgcctggc	tgaccgcccc	acgacccccc	cccatgtacg	tcaataatga	cgtatgttcc	480
catagttaacg	ccaaatggga	ttttccattt	acgtcaatgg	gtggagttt	tacgttaaac	540
tgcccacttg	gcagttacatc	aagtgtatca	tatgcctaaat	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggccgcct	ggcatttatgc	ccagttacatg	actttatggg	actttcttac	660
ttggcagttac	atctacgtat	tagtcatcg	tattaccatg	gtgatgcggg	tttggcagta	720
catcaatggg	cgtggatagc	gttttgactc	acggggattt	ccaagtctcc	accccccattga	780
cgtcaatggg	agttttgtttt	ggcacccaaa	tcaacgggc	tttccaaaat	gtcgttaacaa	840
cgtccggccca	tttgacgaaa	tggggcggtt	gcgtgtacgg	tgggggttct	atataaggcag	900
agctcgat	gtgaaacggc	agatgcctg	gagacccat	ccacgcgtt	ttgacccatcc	960
tagaagacac	cgggaccgat	ccagccctcc	cgccggggaa	cgtgcatttgc	gaacgcggat	1020
tcccccgtgcc	aagagtgcac	taagtaccgc	ctataactc	tataggcaca	ccccctttggc	1080
tcttatgcat	gctatactgt	ttttggctt	gggcctatac	accccccgtt	ccttatgtct	1140
taggtgtatgg	tatagcttag	cctataggtt	tgggttatttgc	accattatttgc	accactcccc	1200
tattttgttgc	gatactttcc	attactaattc	cataacatgg	ctctttgc	caactatctc	1260
tattttgttgc	atgccaataac	tctgtccccc	agagactgac	acggactctg	tattttttaca	1320
ggatgggggtc	ccatttattt	tttacaaaattt	cataatatac	acaaacggcg	cccccggttgc	1380
cgcaatgtttt	attaaacata	ggctgggatc	tccacgcga	tctcggttgc	gtgtccggaa	1440
catgggcctt	tctccggtag	ggggggagct	tccacatcc	agccctggc	ccatgcctcc	1500
agcggcgtcat	ggtcgtcgg	cagtccttgc	ctcttaacag	tggaggccag	acttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtccgc	cacaaggccg	tggccgttgc	gtatgtgttgc	1620
gaaaatgagc	gtggagatttgc	ggctcgacgc	gctgcacgc	atgaaagact	taaggcagcg	1680
gcagaagaag	atgcaggcag	ctgagtttttgc	gtatttctgt	aagagtgcag	gttactccc	1740
gttgcgggtc	tgtaaacgg	ggggggcagtt	gtagtcgt	cagttactcg	tgtgtccgg	1800
cgccaccca	gacataatag	ctgacagact	aacagactgt	tcctttccat	gggtcttttc	1860
tgcagtccacc	gtccttagat	caccatggat	gcaatgaaga	gagggtctgc	ctgtgtgttgc	1920
ctgctgtgt	gagcagtctt	cgtttgc	agcgagatct	gctgtgcctt	ctagttgc	1980
gcatctgtt	gtttggccctt	cccccggttgc	ttccttgacc	ctggaaagggt	ccactccac	2040
tgtcccttttgc	taataaaaatgc	ggggaaatttgc	atcgcatttgc	ctgagttttgc	gtcatttttat	2100
tctgggggggt	gggggggggc	aggacagca	ggggggaggat	tggaaagaca	atagcaggca	2160
tgctggggat	gccccgggttgc	ctatggccgc	tgcggccagg	ttgtgttgc	ttgacccgg	2220
tcctcttgc	ccagaaaagaa	gcaggccat	cccccttcttgc	gtgacacacc	ctgtccacgc	2280
ccctgggttgc	tagttccagc	cccactcata	ggacactcat	agctcaggag	ggctccggc	2340
tcaatccccac	ccgcataaagt	acttggagcg	gtctctccct	ccctcatcg	cccacaaac	2400
caaaccatgc	ctccaaagat	ggggaaaat	taaagoaaga	taggttata	agtgcagagg	2460
gagaaaaat	gccttccaa	tgtggaaat	taatggaga	aatcatagaa	tttcttccgc	2520
ttcctcgctc	actgactcgc	tgccgtcggt	cggtcgctg	cggcgagccg	tatcagctca	2580
ctcaaaaggcg	gtaataacgg	tatccacaga	atcaggggat	aacccggaa	agaacatgt	2640
agcaaaaaggc	cagaaaaagg	ccagggaccc	taaaaaggcc	cggttgc	cggttttcca	2700
taggtccccc	ccccctgtac	agcatcaca	aaatcgtacgc	tcaagtgc	ggggccggaaa	2760
cccgacacca	ctataaaagat	accaggcg	tcccccttgc	agctccctcg	tgcgtctcc	2820
tgttccggacc	ctgcccgtt	ccggatacc	gtccgcctt	ctcccttgc	gaagcgttgc	2880
gctttctcat	agtcacgt	gttaggtatc	cagttccgtt	taggtcg	gctccaaatcg	2940
gggctgtgt	cacgaaacccc	ccgttcacgc	cgaccgcgt	gccttaccc	gtaactatcg	3000
tctttagtcc	aaacccggta	gacacgactt	atcgcact	cgacgcacca	ctggtaacag	3060
gatttagcaga	gcggaggtat	taggcgttgc	tacagatgc	ttgaaagggt	ggccataacta	3120
cggtctacact	agaagaacag	tatgggtat	ctgcgtctg	ctgaagcc	ttaccccttgc	3180
aaaaagagg	gttagtctt	gatccggca	acaaaccacc	gctggtagc	gtggtttttt	3240
tgtttgc	cacgacat	cgcccgat	aaaaggat	caagaagatc	ctttgtatcc	3300
ttctacgggg	tctgcgtc	agttggaaac	aaactc	taaggat	tggtcatgag	3360
attatcaaaa	aggatcttca	cctagatct	ttttaat	aaatgaat	ttaaatcaat	3420
ctaaagtata	tatgat	tttgc	cagttac	tggttata	gtgaggacc	3480
tatctcagcg	atctgtctat	ttcggttcatc	catagttgc	tgactccgg	ggggggggcc	3540
ctgagggtcg	cctcgta	aggtgttgc	tgactcatac	caggccttgc	tcgccccatc	3600
atccagccag	aaagtgggg	agccacgg	gatggagat	ttgttgc	tggaccat	3660
ggtgat	aacttttgc	ttgcccacgg	acgggtcg	ttgtcgggaa	gatgcgtgat	3720
ctgtatcc	aactcag	aaatgtcg	tttcaacaa	agccggcg	ccgtcaagtc	3780

agcgtaatgc tctgccagtgttacaaccaa ttaaccaatt ctgatttagaa aaactcatcg	3840
agcatcaaat gaaactgcaa ttatttcata tcaggattat caataccata ttttgaaaaa	3900
agccgtttct gtaatgaagg agaaaactca ccgaggcagt tcacatggat ggcaagatcc	3960
tggtatcggt ctgcgattcc gactcggtcc acatcaatac aacctttaa ttcccctcg	4020
tcaaaaataa ggttatcaag tgagaaatca ccatgagtga cgactgaatc cggtgagaat	4080
ggcaaaaact tatgcatttc ttccagact tggtaaacag gcagccatt acgctgtca	4140
tcaaaaatcac tcgcataac caaaccgtta ttcatcggt attgcgcctg agcgagacga	4200
aatacgcgat cgctgtaaa aggacaatta caaacaggaa tcgaatgcaa ccggcgcagg	4260
aacactgcca ggcgcataac aatattttca cctgaatcgat gatattcttc taataacctgg	4320
aatgctgtt tccccgggat cgcaagtggtg agtaaccatg catcatcagg agtacggata	4380
aaatgcttga tggctggaaagg aggcataaat tccgtcagcc agtttagtct gaccatctca	4440
tctgttacat cattggcaac gtcacccctt ccattgttca gaaacaactc tggcgcatcg	4500
ggcttcccat acaatcgata gattgtcgca cctgatttgc cgacattatc gcgagccat	4560
ttatacccat ataaatcagc atccatgttg gaatttaatc gcggcctcga gcaagacgtt	4620
tcccgttgaat tattgtcat aacaccctt gtattactgt ttatgttaagc agacagttt	4680
attgttcatg atgatatatt ttatcttgc gcaatgttac atcagagatt ttgagacaca	4740
acgtggctt ccccccccccc cattattga agcatttac aggttatttgc tctcatgagc	4800
ggatacatat ttgaatgtat tttagaaaaat aaacaatag gggttccgcg cacatttccc	4860
cggaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaaat	4920
aggcgtatca cgaggccctt tcgtc	4945

<210> 10
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide

<400> 10
ctatataagc agagctcggt tag

23

<210> 11
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide

<400> 11
gtagcaaaga tcttaaggacg gtgactgcag

30

<210> 12
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide

<400> 12
gtatgtgtct gaaaatgagc gtggagatttgc ggctcgac

39

<210> 13
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide

<400> 13

gtgcgagccc aatctccacg ctcatttca gacacatac

39

<210> 14

<211> 4432

<212> DNA

<213> E. coli (V1J plasmid)

<400> 14

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gaagctcccg	gagacggta	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcaggcgcg	tcagcgggt	120
ttggcgggtg	tcggggctgg	cttaactatg	cgccatcaga	gcagattgt	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgttaaggag	aaaataccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgtatccata	tcataaatg	tacatttata	ttggctcatg	300
tccaaacatta	ccgcgtatgt	gacattgtt	attgactagt	tattaatagt	aatcaattac	360
ggggtcattt	gttcatagcc	catatatgga	gttccgcgtt	acataactta	cggtaatagg	420
cccgccctggc	tgaccggcca	acgacccccc	cccatggacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaatagggg	ctttccattt	acgtcaatgg	gtggagttt	tacgtaaac	540
tgcccacttg	gcagttacatc	aagtgtatca	tatgccaatg	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggccgcct	ggcattatg	ccagttacatg	actttatggg	actttccatc	660
ttggcgtatc	atctacgtat	tagtcatcg	tattaccat	gttatgcgg	tttggcgtat	720
catcaatggg	cgttgatgc	gttttgactc	acgggattt	ccaaatctcc	accccatgt	780
cgtcaatggg	agtttggttt	ggcacaaaaa	tcaacggac	tttccaaat	gtcgtaaac	840
ctccggccca	ttgacgcaaa	tggggcgtat	gcgtgtacgg	ttggagggt	tataaagcag	900
agctcggtt	gtgaaccgtc	agatgcctg	gagacccat	ccacgcgtt	ttgaccttca	960
tagaagacac	cgggaccgat	ccagcctccg	cgccgggaa	cgtgtcattt	gaacgcggat	1020
tccccgtgcc	aagagtgcg	taagtaccgc	ctatacgatc	tataggccca	cccccttggc	1080
ttcttatgtt	tgctatactg	tttttggctt	ggggcttata	caccccccgt	tcctcatgtt	1140
ataggtgtat	gtatacgat	gttatacgat	gtgggttatt	gaccattatt	gaccactccc	1200
ctattgggt	cgatacttgc	cattactaat	ccataacatg	gcttttgc	acaactctt	1260
ttattggcta	tatgccaata	cactgtcctt	cagagactga	cacggactct	gtatttttac	1320
aggatgggtt	ctcattttt	atttacaaat	tcacatatac	aacaccaccg	tccccagtgc	1380
ccgcagttt	tattaaacat	aacgtggat	ctccacgcga	atctgggtt	cgtgtccgg	1440
acatgggtt	ttctccgtt	ggggcggagc	ttctacatcc	gagccctgt	cccatgcctc	1500
cagcgactca	tggtgcgtc	cgacgtcc	gctctttaaca	gtggaggc	gacttaggc	1560
cagcactatg	cccaccacca	ccagtgtcc	gcacaaggcc	gtggcggtag	gttatgttgc	1620
tgaaaatggag	ctcggggagc	gggcttgcac	cgctgacgc	tttggaaagac	ttaaggcagc	1680
ggcagaagaa	gatgcaggca	gctgagttt	tgtgttctga	taagagtca	aggttaactcc	1740
cgttgcgtt	ctgttaacgg	tgaggggcag	tgttagtctga	gcagtactcg	ttgctgcgc	1800
gcccgcacc	agacataata	gctgacagac	taacagactg	ttcccttcca	ttggcttttt	1860
ctgcagtac	cgtcctttaga	tctgctgtgc	cttctagtt	ccagccatct	tttggtttgc	1920
cctcccccgt	gccttcctt	accctggaa	gtgccatcc	cactgtcctt	tcctaataaa	1980
atgagggaaat	tgcacatcgat	tgtctggat	gggtgttatt	tattctgggg	gttgggggtgg	2040
ggcagcacag	caagggggag	gattgggaa	acaatagcag	gcatgctggg	gatgcgggtt	2100
gctctatggg	tacccagggt	ctgaagaatt	gaccgggtt	ctccctggcc	agaaaagaac	2160
aggcacatcc	ccttcctctgt	gacacaccct	gtccacgccc	ctgggttctt	gttccagccc	2220
cactcatagg	acactcatag	ctcaggaggg	ctccgcctt	aatcccaccc	gctaaagatc	2280
tggagcgtt	cttccttcc	ctcatcagcc	caccaaaacca	aacctagctt	ccaagagtgg	2340
gaagaaat	aagcaagata	ggctttaa	tgcagggag	gagaaaatgc	ctccaaatcg	2400
tgaggaagta	atgagagaaa	tctatagaatt	tcttccgtt	cctcgcttac	tgactcgct	2460
cgtctgggtc	ttcggtgtcg	ggcggcggt	tcaatccact	caaaggccgt	aatacggtt	2520
tccacagaat	caggggataa	cgcaggaaag	aacatgtgag	caaaaggcca	gaaaaaggcc	2580
aggaaccgtt	aaaaggccgc	gttgcgttgc	ttttccata	ggctccgccc	ccctgacgag	2640
catcacaaaa	atcgacgc	aaatcgaggg	tggcgaacc	cgacaggact	ataaaagatc	2700
caggcgttcc	ccccctggaa	ctccctcggt	cgcttcctt	ttccgcaccc	gcccgttacc	2760
ggatacctgt	ccgccttct	cccttcggga	agcgtggc	tttctcaatg	ctcacgcgt	2820
aggtatctca	gttcgggtt	gttcgttgc	tccaagctt	gctgtgttgc	cgaaacccccc	2880
gttca	ccgcgttgc	ttatccgtt	aactatcg	tttgcgttccaa	cccggttaca	2940
cacgacttat	cgccactggc	agcagccact	ggttacacgg	ttagcagac	gaggatgtt	3000
ggcgggtgtt	cagaggctt	gaagtgggt	ccttactacg	gctacactag	aaggacagta	3060
tttggtatct	gcgcgttgc	gaagccagg	accttccgaa	aaagagtgg	tagcttctga	3120
tccggcaaac	aaaccacccgc	ttgttagccgt	ggttttttt	tttgcaagca	gcagattacg	3180

cgcagaaaaa	aaggatctca	agaagatcc	ttgatcttt	ctacgggtc	tgacgctcg	3240
tggAACGAAA	actcacgta	agggatTTT	gtcatgagat	tatcaaaaAG	gatcttacc	3300
tagatcTTT	taaattaaaa	atgaagTTT	aatcaatct	aaagtata	tgagtaaact	3360
tggTCTgaca	gttaccaatg	cTTaatcagt	gaggcaccta	tctcagcgt	ctgtctattt	3420
cgttcatcca	tagttgcctg	actccccgtc	gtgtagataa	ctacgatacg	ggagggtta	3480
ccatctggcc	ccagtctgc	aatgataccg	cgagacccac	gctcaccggc	tccagattt	3540
tcagcaataa	accagccagc	cggaagggcc	gagcgcagaa	gtggctc	aactttatcc	3600
gcctccatcc	agtctttaa	ttgttgcgg	gaagctagag	taagttagt	gccagttat	3660
agtttgcgca	acgttgttgc	cattgtcata	ggcatcg	tgcacgtc	gtcgttgg	3720
atggcttcat	tcagtcgg	ttcccaacga	tcaaggcg	ttacatgatc	ccccatgtt	3780
tgcaaaaaag	cggtagtgc	ttcggtcc	ccgatcg	tcaagaat	gttggccgca	3840
gtgttatcac	tcatggttat	ggcagactg	cataatttctc	ttactgtcat	gccatccgt	3900
agatgcttt	ctgtgactgg	ttagtactca	accaagtcat	tctgagaata	gtgtatgcgg	3960
cgaccgagtt	gctcttgc	ggcgtcaata	cgggataata	ccgcgc	tagcagaact	4020
ttaaaagtgc	tcatcatttg	aaaacgttct	tcggggcgaa	aactctcaag	gatcttaccg	4080
ctgttgat	ccagtccat	gtacccact	cgtgcaccc	actgtatcc	agcatcttt	4140
actttcacca	gcgttctgg	gtgagaaata	acagggaggc	aaaatgcgc	aaaaaaggga	4200
ataagggcga	cacggaaatg	tttataactt	atactttcc	ttttcaata	ttattgaagc	4260
atttatcagg	gttattgtct	catgagcg	tacatattt	aatgtat	aaaaataaaa	4320
caaatagggg	ttccgcgcac	atttccccga	aaagtgcac	ctgacgtct	agaaaccatt	4380
attatcatga	cattaaccta	taaaaaatagg	cgtatcacga	ggccctttcg	tc	4432

<210> 15

<211> 4864

<212> DNA

<213> E. coli (V1Jneo plasmid)

<400> 15

tcgcgcgtt	cggtgatgac	ggtaaaaacc	tctgacacat	gcagctcc	gagacggta	60
cagctgtct	gtaaaggcgat	ggccggagca	gacaagcc	tcaggcgc	tcagcgggt	120
ttggcgggt	tcggggctgg	cttaactat	cggtcata	gcagatgt	ctgagatgc	180
acatatgcg	gtgtgaaaata	ccgcacagat	cgtaaggag	aaaataccgc	atcagattt	240
ctattggca	ttgcatacgt	tgtatccata	tcataatat	tacattata	ttggctcat	300
tccaacatta	ccgcatgtt	gacattgatt	attgactat	tattaatagt	aatcaattac	360
ggggtcattt	gttcatagcc	catatatgg	gttccgc	acataactt	cggttaatgg	420
cccccctggc	tgaccgc	acgaccc	cccattgac	tcaataat	cgtatgttcc	480
catagtaacg	ccaaatagg	cttccattt	acgtcaat	gtggagtatt	tacggtaaac	540
tgcccaactt	gcagtacatc	aagtgtat	tatgc	acgc	ttgacgtca	600
tgacggtaaa	tggccgc	ggcattat	ccgtacat	acccat	actttccat	660
ttggcgtac	atctacgtat	tagtcat	tattaccat	gtgatgc	tttggcgt	720
catcaatgg	cgtggat	ggtttact	acgggat	ccaagtct	accccat	780
cgtcaatgg	agtttgg	ggcacaaa	tcaacgg	tttccaaa	gtcgtaacaa	840
ctccgc	ttgaccaaa	tggcggt	gcgtgtac	tggagg	atataagcag	900
agctcg	gtgaaaccgt	agatcg	gagaccc	ccacgct	ttgaccttca	960
tagaagac	cgggacc	ccagcc	cgccgg	cggtgc	gaaacgcgg	1020
tcccg	aaaggtgac	taagtacc	ctatagat	tatagg	cccccttgg	1080
ttcttat	tgctata	ttttgg	gggtctata	cacccc	tcctcat	1140
atagggtat	gtata	gcttat	gtgggtt	gaccattt	gaccact	1200
ctattgg	cgata	cattact	ccataacat	gctttt	acaactct	1260
ttattgg	tatg	cactgt	cagagact	gacggact	gtat	1320
aggatgg	ctcattt	atttacaa	tcacata	aacaccac	ccccag	1380
ccgcag	tat	aaac	acgtgg	ctccac	cggttcc	1440
acatgg	ttctcc	ggcgg	ttctacat	gagcc	cccatgc	1500
cagc	tggtc	gcag	tcctt	gctc	gactt	1560
cagc	cccac	ccag	gtgc	gcaca	ggatgt	1620
tggaa	ctcg	ggg	tcgt	gtgg	ttaa	1680
ggc	gag	gtcg	gac	ggag	aggtaact	1740
cggt	ctgtt	ttgt	gtgtt	ttaa	gttgc	1800
gcgc	acata	gtcg	actgt	ggat	ttggctt	1860
ctgc	ctgc	tctg	gttct	ttgg	ttgttgc	1920
cctcc	ccgt	ccat	ccag	ccat	ttctaa	1980
atgagg	at	tgcat	gtgt	tat	tttgg	2040

ggcagcacacg	caaggggggag	gattgggaag	acaatagcag	gcatgctggg	gatgcgggtgg	2100
gctctatggg	taccagggt	ctgaagaatt	gaccgggttc	ctccctggggc	agaaaagaagc	2160
aggcacatcc	ccttctctgt	gacacacccct	gtccacgccc	ctggttctta	gttccagccc	2220
cactcatagg	acactcatag	ctcaggaggg	ctccgccttc	aatcccaccc	gctaagagtac	2280
tttagggcggt	cttcctcc	ctcatcagcc	caccaaacca	aacctagcct	ccaagagtgg	2340
gaagaaatta	aagaagata	ggcttataag	tgcaagggga	gagaaaaatgc	ctccaaacatg	2400
tgtaggaagta	atagagagaaa	tcatagaat	tcttcgctt	cctcgctcac	tgactcgctg	2460
cgctcggtcg	ttcggtcg	gcgacgggt	tcagtcact	caaaggcggt	aatacggtta	2520
tccacagaat	caggggataa	cgcagggaaag	aacatgtgag	aaaaaggcga	gaaaaaggcc	2580
aggaaccgt	aaaaggccgc	gttgctggcg	tttttcata	ggctccgccc	ccctgacgag	2640
catcacaaaa	atcgacgctc	aagttagagg	tggcgaaacc	cgacaggact	ataaaagatac	2700
caggcgttt	ccccctggaa	ctccctcg	cgctctctg	tccgaccct	gcccgttacc	2760
ggataccatgt	ccgccttct	cccttcggga	agcgtggcgc	tttctcaatg	ctcacgtgt	2820
aggtatctca	gttcgggt	gttgcgttgc	tccaaagtgg	gttgtgtgc	cgaaccccccc	2880
gttcagcccc	accgtgcgc	tcttacccgt	aactatcg	ttgagtccaa	cccggttaaga	2940
cacgacttat	cgccactggc	agcagccact	ggtaaaggg	ttagcagagc	gaggtatgt	3000
ggcggtgcta	cagatttctt	gaagtgggtt	cctaactacg	gtcatactag	aggacagta	3060
tttggtatct	gchgctctgt	gaagccagtt	accttcggaa	aaagagttgg	tagcttga	3120
tccggcaaac	aaaccaccgc	tggtagcggt	gtttttttt	tttgcagca	gcagattacg	3180
cgcagaaaaa	aaggatctca	agaagatct	ttgatctttt	ctacggggtc	tgacgctcag	3240
tggaaacggaa	actcacgtt	agggattttt	gtcatagat	tatcaaaaag	gatcttcacc	3300
tagatccctt	taattaaaaa	atgaagttt	aaatcaatct	aaagtatata	tgagtaaact	3360
tggtctgaca	gttaccaatg	cttaatca	gaggccacta	tctcagcgat	ctgtctattt	3420
cgttcatcca	tagttgc	actccggggg	ggggggggcgc	tgagggttgc	ctcgtaaga	3480
aggtgttgc	gactcatacc	aggcctgaat	cgccccatca	tccagccaga	aagtggggga	3540
gccacgggt	atgagagctt	tgtttaggt	ggaccagtt	gtgattttga	acttttctt	3600
tgccacggaa	cgggtcg	tgtcgaaag	atgcgtgatc	tgatccttca	actcagcaaa	3660
agttcgat	atcaacaaaa	gcccgtcc	cgtcaagtca	gcgttaatgc	ctggcagtgt	3720
tacaaccaat	taaccaattc	tgattagaaa	aactcatcga	gcataaattg	aaactgcaat	3780
ttattcatat	caggattatc	aataccat	ttttaaaaaa	ggcgtttctg	taatgaagga	3840
aaaaactcac	cgaggcagtt	ccataggat	gcaagatcct	gttatcg	tgcgatccg	3900
actcgtccaa	catcaataca	acctattaa	ttccccctcg	aaaaataaaag	gttatacaat	3960
gagaatcac	catgagt	gactgaatcc	ggtgagaatg	gaaaaagctt	atgcatttct	4020
ttccagactt	gttacacagg	ccagccat	cgctcgat	aaaaatcact	cgcatcaacc	4080
aaaccgttat	tcatcg	ttgcgc	gcgagacgaa	atacgcgatc	gctgttaaaa	4140
ggacaattac	aaacaggaat	cgatgc	cgccgcagg	acactgccc	cgcatcaaca	4200
atattttcac	ctgatcagg	atattttc	aataactgg	atgtctttt	cccggggatc	4260
gcagtggta	gtacccatgc	atcatcagg	gtacggataa	aatgttgc	gttgcggaa	4320
ggcataaatt	ccgtcagcc	gttttagt	accatctcat	ctgtacatc	attggcaacg	4380
ctacccttgc	catgtt	aaacaactct	ggcgcatcg	gttcccata	caatcgatag	4440
attgtcgac	ctgatt	gacattatcg	cgagcccatt	tataccata	taaatcagca	4500
tccatgttgg	aatttaatcg	cgccctcg	caagacgtt	cccggttgaat	atggctcata	4560
acaccccttg	tattactgtt	tatgtaa	gacagtttta	ttgttcatga	tgatataattt	4620
ttatcttgc	caatgtaa	tcagat	tgagacacaa	cgtggcttcc	cccccccccc	4680
cattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	4740
tagaaaaata	aacaaatagg	ggttccgc	acattcccc	gaaaagtgc	acctgacgtc	4800
taagaaacca	ttattatcat	gacattaacc	tataaaaata	ggcgtatc	gaggccctt	4860
cgtc						4864

<210> 16
<211> 4867
<212> DNA
<213> E. coli (V1Jns plasmid)

cccgccctggc	tgaccgcccc	acgacccccc	cccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggga	cttccatttgc	acgtcaatgg	gtggagtatt	tacggtaaac	540
tgcccacttgc	gcagtacatc	aagtgtatca	tatgccaatgt	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggccgccttgc	ggcatttatgc	ccagttacatg	accttatggg	actttcttac	660
ttggcgtatc	atctacgtat	tagtcatcg	tattaccatg	gtgatgcgtt	tttggcgtat	720
catcaatggg	cgtggatagc	ggtttactc	acggggattt	ccaagtctcc	accccattga	780
cgtcaatggg	agtttggttt	ggcacaaaaa	tcaacgggac	tttccaaaat	gtcgtaaaca	840
ctccgcccc	ttgaccaaa	tgggcgttag	gcgtgtacgg	tgggaggctt	atataagcag	900
agctcggttta	gtgaaccgtc	agatgcctgt	gagacgccc	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgt	ccagccccc	cggccggaa	cgtgttttttgc	gaacgcggat	1020
tccccgtgcc	aagatgtacg	taagtaccgc	ctatagactc	tataggcaca	ccccctttggc	1080
tctttagatgc	gctataatgt	ttttggcttgc	gggccttatac	accccgttgc	ccttatgtct	1140
taggtgtatgg	tatactttag	cctataggttgc	tgggttatttgc	accattatttgc	accactcccc	1200
tattgggtatgc	gatactttcc	attactaattc	cataacatgg	ctctttggca	caactatctc	1260
tattggcttatgc	atgccaataac	tctgtccttc	agagactgac	acggactctg	tatttttaca	1320
ggatgggggtc	ccatttatttgc	tttacaaaatttgc	cacatataca	acaacgcctgt	cccccggtgcc	1380
cgccgttttttgc	attaaacata	gcgtgggatc	tccacgcgaa	tctcggttgc	gtgttccggaa	1440
catgggtcttgc	tctccggtag	cgccggagcttgc	tccacatccgc	agccgttgc	ccatgcctcc	1500
agcggctcatgc	gttcgtctgg	cagtccttgc	ctcctaacag	tggaggccag	actttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtggcc	cacaaggccc	tggcggttagg	gtatgtgtct	1620
gaaaatgagc	gtggagatttgc	ggctcgacgc	gctgacccgc	atggaaagact	taaggcagcg	1680
gcagaagaag	atgcaggcag	ctgagttgttgc	gtattctgtat	aagagtcaga	gttaactccc	1740
gttgcgtgtc	tgttaacggt	ggagggcagt	gtagtctgt	cagtaactcgt	tgctgcccgg	1800
cgccgcccc	gacataatag	ctgacagact	aacagactgt	tctttccatgc	gggtcttttc	1860
tgcaatgc	gtccatgtat	ctgtgtgtcc	ttcttagttgc	cagccatcttgc	tttggggggccc	1920
ctcccccgttgc	catttcatttgc	cccttggaaagg	tgccatcccc	actgtcccttgc	cctaataaaaa	1980
tgaggaaatttgc	gcatcgatttgc	gtctgagtag	gtgttatttgc	attctgggggg	gtgggggtggg	2040
gcaggacacgc	aaggggggagg	atttggaaaga	caatagcagg	catgtgtgggg	atgcgggtggg	2100
ctctatggcc	gctggggcca	ggtgctgaag	aatttgcaccgc	gttcccttgc	ggccagaaaag	2160
aaggcaggcac	atcccttgc	ctgtgacaca	ccctgtccac	gccccctgggtt	cttagttcca	2220
gccccactca	taggacactc	atagctcagg	agggctccgc	cttcaatccc	acccgcataaa	2280
gtacttggag	cggtctctcc	ctcccttcatc	agcccaacca	accaaactca	gcctccaaga	2340
gtgggaaaggaa	attaaagcaa	gataggctatgc	taatgtcaga	gggagagaaa	atgcctccaa	2400
catgtgagaa	agtaatgaga	gaaatcatag	aatttttcc	gttcccttgc	tcaactgactc	2460
gctgcgtctgc	gtcggttgc	tgccggcgacgc	ggtatcagct	cactcaagg	cggtatatacg	2520
gttatccaca	gaatcagggg	ataacgcagg	aaagaacatgc	tgagcaaaa	gccagcaaaa	2580
ggccaggaaac	cgtaaaaaagg	ccgcgttgc	ggcggttttgc	cataggctcc	cccccccttgc	2640
cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aacccgacag	gactataaaag	2700
ataccaggcg	tttcccccttgc	gaagctcccttgc	cgtgcgttgc	ccctgtccgc	ccctgtccgc	2760
taccggatac	ctgtccgc	tttcccttgc	gggaaagctgt	gcttttctc	atagctcagc	2820
ctgttagtatgc	ctcaggatgc	tgttaggtatgc	tcgttccaa	cttgggtgt	tgcacgaacc	2880
ccccgttccatgc	cccgaccgc	gccccttattc	cggttaactat	cgtttgcgtat	ccaaacccgg	2940
aagacacacgc	ttatcgccac	ttgcagcaggc	cactggtaac	aggattagca	gagcgaggtt	3000
tgttaggcgttgc	gctacagatgc	tcttgcgtatgc	gtggccatca	tacggctaca	ctagaagaac	3060
agtattttgttgc	atctgcgttgc	tgctgaaatgc	agttaccc	ggaaaaaagag	ttggtagctc	3120
ttgatccgc	aaacaaacca	ccgcgttgc	cggtgggttttgc	tttgcgtatgc	agcagcagat	3180
tacgcgcaga	aaaaaaaggat	tccaaagaa	tcctttgtatgc	ttttcttgc	ggtctgcacgc	3240
tcagtggaaatgc	aaaaacttacac	tttgcgtatgc	ttttgttgc	ttttgttgc	aaaggatctt	3300
caccttagatc	tttttaaaatttgc	aaaaatgaaatgc	ttttaaatca	atctaaatgc	tatatgagta	3360
aacttggtct	gacagttacc	aatgtcttac	cagtggaggca	cctatctc	cgatctgtct	3420
atttcgttca	tccatagttgc	cctgactcgg	gggggggggg	cggtgggttgc	tgcctgtatgc	3480
agaaggtgttgc	gctgactcat	accaggcctgt	aatcgccccca	tcatccagcc	agaaagttag	3540
ggagccacgg	tttgcgtatgc	tttgcgtatgc	tttgcgtatgc	tttgcgtatgc	tttgcgtatgc	3600
ctttggccacgc	gaacgttgc	cggtgggttgc	aaagatgttgc	atctgtatgc	tcaactcagc	3660
aaaaggatgc	tttgcgtatgc	aaagcccccgc	tttgcgtatgc	tttgcgtatgc	tttgcgtatgc	3720
tgttacaaccatgc	aaatcaaccatgc	tttgcgtatgc	aaaaactcat	cgagcatca	atgaaactgc	3780
aattttatgc	tatcaggatgc	atcaataccatgc	tatttttgc	aaagccgtttgc	ctgtatgc	3840
ggagaaaaacttgc	caccggaggca	gttccataggc	atggcaagat	cctggatgc	gtctgcgttgc	3900
ccgactcgatc	caacatcaatgc	acaaccttgc	aattttcccttgc	cgtaaaaaat	aagggttatca	3960
agtggaaaatgc	caccatgaggatgc	gacgactgtatgc	tccgggtgaga	atggcaaaa	ctttagtgc	4020
tctttccatgc	tttgcgtatgc	aggccaggca	ttacgcgtatgc	catcaaaaatgc	actcgcatca	4080
acccaaacccgttgc	tattcattgc	tgatttgcgtatgc	tgagcgagac	gaaatacgc	atcgctgttgc	4140

aaaggacaat tacaaaacagg aatcgaatgc aaccggcgca ggaacactgc cagcgcatca	4200
acaatatttt cacctgaatc aggatattct tctaataccct ggaatgctgt ttccccgggg	4260
atcgcagtgg tgagtaacca tgcatacatca ggagtacgga taaaatgctt gatggtcgga	4320
agaggcataa attccgttag ccagtttagt ctgaccatct catctgtaac atcattggca	4380
acgctacctt tgccatgttt cagaaacaac tctggcgcat cgggcttccc atacaatcga	4440
tagattgtcg cacctgattt cccgacatca tcgcgagccc attataccca atataaatca	4500
gcacccatgt tggaatttaa tcgcggccctc gagcaagacg ttcccggttg aatatggctc	4560
ataacacccc ttgttattact gtttatgtaa gcagacagtt ttatgttca tgatgatata	4620
tttttatctt gtgcataatgtt acatcagaga ttttgagaca caacgtggct ttccccccc	4680
ccccattattt gaagcattta tcagggttat tgtctcatga gggatacat atttgaatgt	4740
attttagaaaa ataaaacaaat aggggttccg cgcacatttc cccgaaaagt gccacctgac	4800
gtctaagaaa ccattattat catgacatta acctataaaa ataggcgtat cacgaggccc	4860
tttcgtc	4867
<210> 17	
<211> 75	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 17	
gatcaccatg gatgcataatga agagagggtctgtgtgtg gaggcgttcc	60
cgtttcgccc agcga	75
<210> 18	
<211> 78	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 18	
gatctcgctg ggcgaaaacga agactgctcc acacagcagc agcacacagc agagccctct	60
cttcattgca tccatgtt	78
<210> 19	
<211> 33	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 19	
ggtacaaaata ttggctattt ggcatttgcac acg	33
<210> 20	
<211> 36	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 20	
ccacatctcg aggaaccggg tcaatttcttc agcacc	36
<210> 21	
<211> 38	

<212> DNA
 <213> Artificial Sequence

 <220>
 <223> oligonucleotide

 <400> 21
 ggtacagata tcggaaaagcc acgttgtgtc tcaaaaatc 38

<210> 22
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> oligonucleotide

 <400> 22
 cacatggatc cgtaatgctc tgccagtgtt acaacc 36

<210> 23
 <211> 39
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> oligonucleotide

 <400> 23
 ggtacatgtat cacgttagaaa agatcaaagg atcttcgg 39

<210> 24
 <211> 35
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> oligonucleotide

 <400> 24
 ccacatgtcg acccgtaaaaa aggccgcgtt gctgg 35

<210> 25
 <211> 4864
 <212> DNA
 <213> E. coli (V1R plasmid)

 <400> 25
 tcgcgcgttt cggtgatgac ggtaaaaacc tctgacacat gcagctcccg gagacggtca 60
 cagcttgctc gtaaggcgat gcccggagca gacaagcccg tcagggcgcg tcagcggtg 120
 ttggcggtt tcggggctgg cttaactatg cggcatcaga gcagattgtt ctgagagtgc 180
 accatatgctc gtgtgaaata cccacacatgc gctaaaggag aaaataccgc atcagattgg 240
 ctattggcca ttgcatacgt tttatccata tcataatatg tacattata ttggctcatg 300
 tccaacatc cccatgtt gacattgtt attgactagt tattaatagt aatcaattac 360
 ggggtcatta gttcatagcc catatatggg gttccgcgtt acataactta cggtaaatgg 420
 cccgcctggc tgaccggcca acgacccccc cccattgacg tcaataatga cgtatgttcc 480
 catagtaacg ccaataggaa ctttccattt acgtcaatgg gtggaggatt tacggtaaac 540
 tgcccacttgc gcaatgtatca tatgcaatgg acgcccccta ttgacgtcaa 600
 tgacggtaaa tggccgcctt ggcattatgc ccagtacatg accttatggg actttccatc 660
 ttggcgttac atctacgtat tagtcatcgc tattaccatg gtgatgcgtt tttggcgttac 720
 catcaatggg cgtggatagc ggtttactc acggggattt ccaagtctcc accccatgaa 780
 cgtcaatggg agtttggggggggcaccaaaa tcaacgggac tttccaaaat gtcgtaaacaa 840

ctccgccccca ttgacgc当地	tggcggtag gcgtgtacgg tgggaggctt atataagcag	900
agctcgat	gtgaaccgtc agatcgctg gagacgccat coacgcgtt ttgacccca	960
tagaagacac	cgggaccgat ccagcctcg cggccggaa cggtcattt gaacgc当地	1020
tccctatcg	aagagtgc当地 taagtaccgc ctataagtc tataggccc ccccttggc	1080
ttcttatcg	tgctatactg ttttggctt ggggtctata caccggcgtt tcctcatgtt	1140
ataggtgt	gtatagtc当地 gcctatagggt gtgggttatt gaccattt gaccactccc	1200
ctattggta	cgatcatcc tacttaat ccataacatg gctcttgc当地 acaactctt	1260
ttattggta	tatgccaata cactgtcctt cagagatgc cagggactt gttatttac	1320
aggatgggt	ctcatttatt atttacaaat tcacatatac aacaccaccg tccccagtgc	1380
cccgagttt	tattaaacat aacgtggat ctccacgc当地 atctgggtt cgttccgg	1440
acatgggtc	ttctccgtt gccggggagc ttctacatcc gagccctgct cccatgc当地	1500
cagcgact	tggcgctcg cgactccctt gtcctaaca gttggaggcc local gacttaggca	1560
cagcacgt	ccccacca ccaatgtgc当地 gcacaaggcc gttggcgtag ggtatgtgc	1620
tgaaaatgg	ctcgccggagc gggcttgc当地 cgtgc当地 ttggaaagac ttaaggc当地	1680
ggcagaagaa	gatgc当地 gctgatgtt tttgttgc当地 taagatgc当地 aggtaaactcc	1740
cgatgc当地	ctgttaacgg tggaggccagc tttgttgc当地 gcagacttgc当地 ttgctccgc	1800
gcgc当地	ccaccaccat gctgacagac taacagactt ttcttttcca tgggtctt	1860
ctgc当地	ctgc当地 tctgttgc当地 ttcttagtgc当地 ccagccatct gttgttgc	1920
cctcccccgt	gcctccgtt accctggaaat gtgc当地 acttgc当地 tcctaataaaa	1980
atgaggaaat	tgc当地 catgc当地 ttgttgc当地 ggtgttcatc tatttctggg ggtgggttgg	2040
ggcagcacag	caagggggag gattggaaatg acaatgc当地 gcatgcttgg gatgc当地	2100
gctctatgg	taccagggtg ctgaaatgc当地 gaccgggtt ctc当地 tggcc	2160
aggcacatcc	ccttctctgt gacacaccctt gtccacgccc ctgggttctt gttccagccc	2220
caactcatgg	acactcatag ctcaaggaggg ctccgc当地 aatcccacc cgtaaagtt	2280
ttggagcggt	ctctccctcc ctcatcagcc caccaaacc aaccttagcct ccaagagtt	2340
gaagaaat	aagcaagata ggctttaatg tgc当地 agggaaatgc当地 ctccaaatgc	2400
tgagaaatg	atgagagaaa tcatgaaatg tcttccgtt ctc当地 gctc当地 tgactc当地	2460
cgctcggtc	ttcgctcg gcgagcggtt tcaatgc当地 caaaggc当地 aatgc当地	2520
tccacagaat	cagggataa cgc当地 agggaaatg aacatgttgc当地 caaaggcc local	2580
aggaaccgta	aaaaggccgc gttgttgc当地 ttttccata ggctccgccc cc当地 tggag	2640
catcacaaaa	atcgacgctc aagtcaagg tggcgaaacc cgacaggact ataaagatac	2700
caggcggtt	ccccctggaaatg ctccctcgatg cgctctctt tttcgaccctt gccgcttacc	2760
ggataccctt	ccgc当地 ttccctggggatg agcgtggccg tttctcaatg ctc当地 cgtt	2820
aggtatctca	gttccgttgc当地 tccaaatgc当地 gttgttgc当地 cgaaccccc	2880
gttcagcccg	accgc当地 gtc当地 cttatccgtt aactatgc当地 ttgatgc当地 cccggtaaga	2940
cacgacttat	cgccacttgc当地 agcagccact ggttacccggatg tttagc当地 gaggtatgt	3000
ggc当地	ggc当地 gtatgttgc当地 gaagtttgc当地 ccttactacg gtttactatg aaggagacta	3060
tttggatatt	gc当地 ctgttgc当地 gaagccaggatg accttccggaa aaagatgttgg tagtcttgc当地	3120
tccggcaaa	aaaccaccgc当地 tggtagccgtt ggttttttgc当地 tttcaagatc gca	3180
cgc当地	aaaggatctca agaagatcc ttgatcttttgc当地 ctatgggggtc tgacgctc当地	3240
tggAACGAAA	actcaatgc当地 agggattttgc当地 gtc当地 agatgatc tatcaaaaatg gatcttacc	3300
tagatccctt	taaattaaaaa atgatgttttgc当地 aatcaatct aatgc当地 tgatcaatc	3360
tggatctgaca	gttaccaatgc当地 cttatcgtt gaggccatc当地 tcttgc当地 ctgtctt	3420
cgatccatcca	tagtgc当地 actccggggatg gggggggccg tgaggctgc当地 ctctgttgc当地	3480
agggtgttgc	gactc当地 accgc当地 ttttgc当地 cggccatc当地 tccaggccatc当地 aagttaggg	3540
gccacgggtt	atgagatgttgc当地 ttgttgc当地 ggaccatgc当地 gtgatgttgc当地 acttttgc当地	3600
tgccacggaa	cggttgc当地 ttgttgc当地 atgc当地 gtatccatc当地 actc当地 gagcaaaa	3660
agttcgattt	attcaacaaa gcccgc当地 cgtcaatgc当地 gctaatgc当地 ctgc当地 cgtt	3720
tacaaccaat	taaccaatgc当地 tgatgttgc当地 aactcatgc当地 gcatcaatgc当地 aactgc当地	3780
ttattcataat	caggattatc aataccatat tttgaaaaatg gccgttctgc当地 taaatgc当地	3840
gaaaactcac	cgaggc当地 ccataggatg gcaagatcc ggtatccgtc当地 tgcatccatc当地	3900
actcgatccaa	catcaataca accttataat ttccctcgatg caaaaataatg gttatcaatc	3960
gagaatcac	catgatgc当地 gactgaatcc ggtgatgc当地 gcaaaatgc当地 atgc当地 atgc当地	4020
ttccagactt	gttcaatgc当地 ccaggccatc当地 cgctcgatc当地 caaatactc cgc当地 catcaacc	4080
aaaccgttat	tcattcgatg ttgc当地 cctgatgc当地 gcgagacccatc当地 atacgc当地 gctgttaaaa	4140
ggacaattac	aaacaggaat cgaatgc当地 cggccatc当地 acactgc当地 cgc当地 catcaaca	4200
atattttcac	ctgaatcagg atattcttc当地 aataccatgc当地 atgc当地 gtttgc当地	4260
gcagtggtg	gtatgc当地 atcatcaggatg gtacggatcc aatgc当地 ggtcgaaaga	4320
ggcataaaatt	ccgtc当地 agggccatc当地 accatctcatc当地 ctgc当地 acatgc当地 atggcaac	4380
ctacccttgc	catgttgc当地 aaacaactctc当地 ggc当地 catccgg gcttccatc当地 caatcgat	4440
attgtcgac	ctgatgttgc当地 gacattatgc当地 cgagccatc当地 tataccatc当地 taaatcgatc	4500
tccatgttgg	aatattaatgc当地 cggccatc当地 caagacgtttt cccgttgc当地 atggctatc	4560

acaccgccttg tattactgtt tatgttaagca gacagttta ttgttcatga tgatatattt	4620
ttatcttgc caatgttaaca tcagagattt tgagacacaa cgtggctttc cccccccccc	4680
cattattgaa gcatttatca gggttattgtt ctcatgagcg gatacatatt tgaatgtatt	4740
tagaaaaata aacaaatagg gttccgcgc acattttccc gaaaagtgcc acctgacgtc	4800
taagaaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt	4860
cgtc	4864
<210> 26	
<211> 36	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 26	
ggtacaagat ctccgcggcc atctcccca ttgaga	36
<210> 27	
<211> 33	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 27	
ccacatagat ctgccccggc tttagtcctc atc	33
<210> 28	
<211> 27	
<212> PRT	
<213> Homo sapien	
<400> 28	
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly	
1 5 10 15	
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser	
20 25	
<210> 29	
<211> 45	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 29	
cagggcgagat ctaccatggc ccccatggc cctattgaga ctgtt	45
<210> 30	
<211> 48	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> oligonucleotide	
<400> 30	
cagggcgagat ctgccccggc tttaatcctc atcctgtcta cttgccac	48

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/34724

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : A61K 48/00; C12Q 1/70.
US CL : 514/44; 435/5; 424/93.1.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 514/44; 435/5; 424/93.1.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Medline, embase, scisearch, biosis, caplus and WEST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	US 6,099,848 A (FRANKEL et al) 08 August 2000 (08.08.2000), page 12 paragraph 6.	1-14, 17
Y	WO 97/31115 A2 (MERCK & CO. INC.), 28 August 1997, page 36.	4
X	WO 90/10230 A1 (UNIVERSITY OF OTTAWA) 07 September 1990, page 11.	17
Y	US 5,858,646 A (KANG) 12 January 1999 (12.01.1999), col. 2, lines 10-17	1-14, 17

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 22 February 2001	Date of mailing of the international search report 09 MAR 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Eleanor Sorbello TERRY J. DEY PARALEGAL SPECIALIST TECHNOLOGY CENTER 1600 Telephone No. 703-308-0196

INTERNATIONAL SEARCH REPORT

Internat	application No.
PCT/US00/34724	

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claim Nos.: 15 & 16
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

THIS PAGE BLANK (USPTO)